

Web site: <http://genome.wustl.edu/gsc>
 Contact: sapiens@wustl.wustl.edu
 Summary Statistics
 Center project name: H_NH0564016

NOTICE: This sequence may not represent the entire insert of this clone. It may be shorter because we only sequence overlapping clone sections once, or longer because we provide a small overlap between neighboring data submissions.

This sequence was finished as follows unless otherwise noted:
 all regions were double stranded, sequenced with an alternate chemistry, or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one subclone; and the assembly was confirmed by restriction digest.

MAPPING INFORMATION:
 Mapping information for this clone was provided by Dr. John D. McPherson, Department of Genetics, Washington University, St. Louis MO. For additional information about the map position of this sequence, see <http://genome.wustl.edu/gsc>

SOURCE INFORMATION:
 The RPCI-11 human BAC library was made from the blood of one male donor, as described by Osoegawa, K., Moon, P.Y., Zhao, B., Frengen, E., Tareno, M., Catanesi, J.U. and de Jong, P.O. (1998) An improved approach for construction of bacterial artificial chromosome libraries. Genomics 51:1-8. The clone may be obtained either from Research Genetics, Inc. (<http://www.resgen.com>) or Pieter de Jong and coworkers at the Roswell Park Cancer Institute (<http://bacpac.med.buffalo.edu>)
 VECTOR: pBAC3.6

NEIGHBORING SEQUENCE INFORMATION:
 The clone sequenced to the left is AC068592; the clone sequenced to the right is CTD-2096A9. Actual start of this clone is at base position 1 of RP11-564016.

The sequence from base position 103517 to 103654 and from base position 103940 to 104887 is derived from PCR of BAC DNA.

The fidelity of the sequence from base position 48371 to 48772 cannot be guaranteed due to an unresolved imperfect dinucleotide repeat.

FEATURES

source

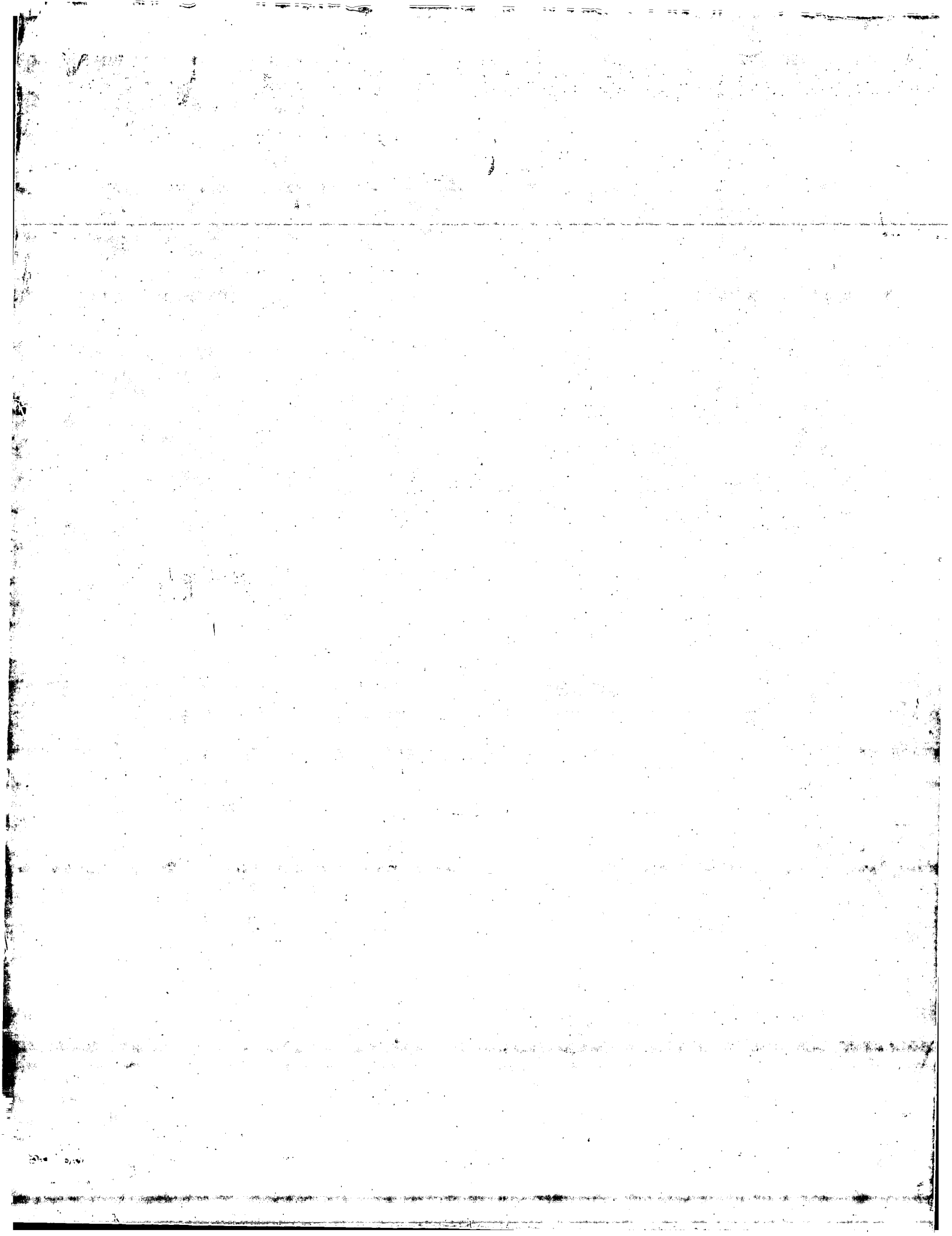
Location/Qualifiers

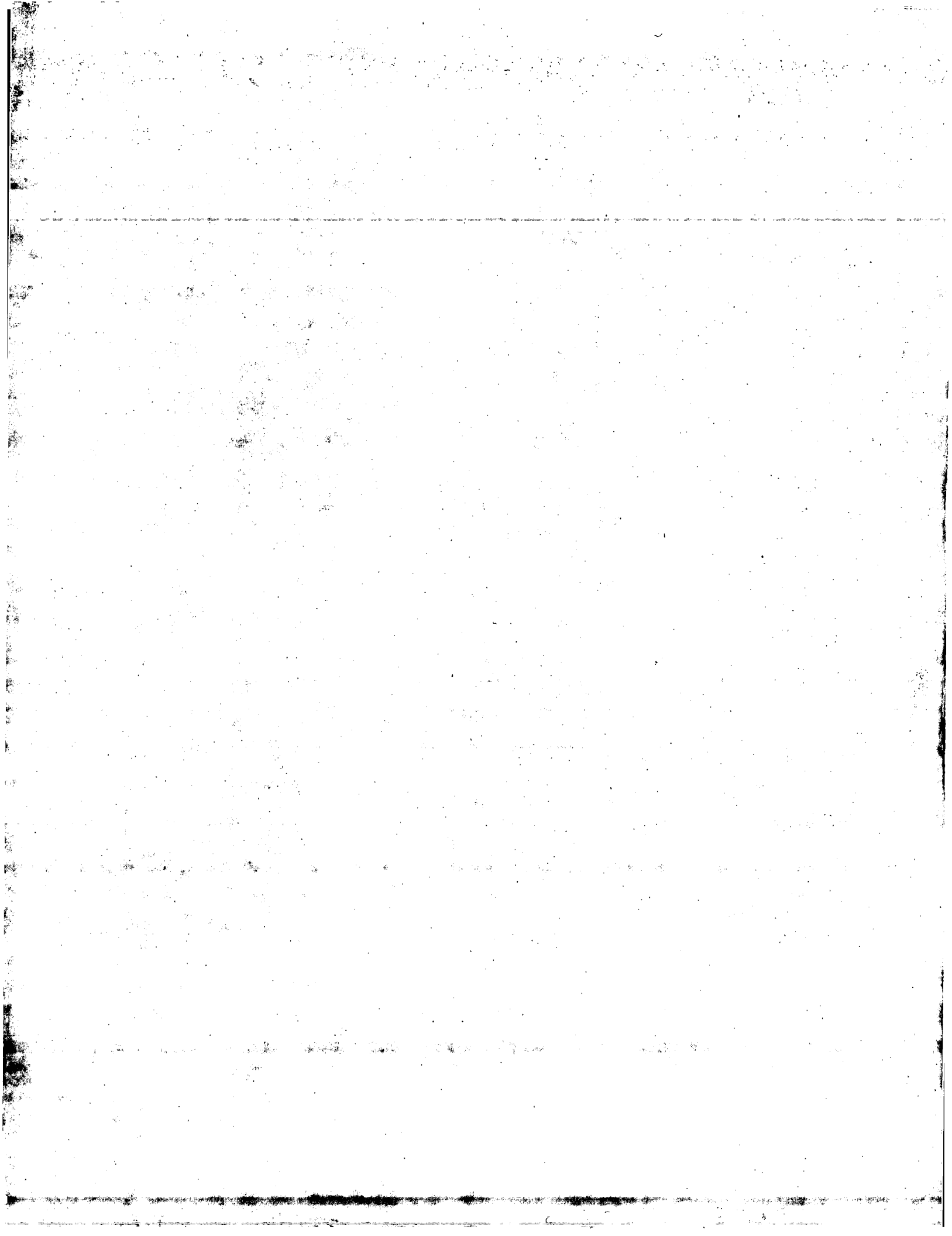
1. 128466
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /chromosome="2"
 /map="2"
 /clone="RP11-564016"
 /clone_id="RPCI-11"
 291..452
 repeat_region
 /rpt_family="Alu"
 1484..1617
 repeat_region
 /rpt_family="L2"
 1669..1826
 repeat_region
 /rpt_family="Alu"
 1864..2412
 repeat_region
 /rpt_family="MIR"
 2413..2591
 repeat_region
 /rpt_family="L2"
 2689..2833
 repeat_region
 /rpt_family="L2"
 3055..3082
 repeat_region
 /rpt_family="L2"
 3085..3392
 repeat_region
 /rpt_family="Alu"
 3350..3643
 repeat_region
 /rpt_family="CR1"
 3644..3717
 repeat_region

repeat_region
 /rpt_family="Mariner"
 3716..4183
 repeat_region
 /rpt_family="CR1"
 6123..6293
 repeat_region
 /rpt_family="MIR"
 6845..6881
 repeat_region
 /rpt_family="AT-rich"
 6911..6963
 repeat_region
 /rpt_family="AT-rich"
 7910..8008
 repeat_region
 /rpt_family="MIR"
 8009..8299
 repeat_region
 /rpt_family="Alu"
 8300..8326
 repeat_region
 /rpt_family="MIR"
 8327..8599
 repeat_region
 /rpt_family="Alu"
 8600..8668
 repeat_region
 /rpt_family="MIR"
 9463..9553
 repeat_region
 /rpt_family="TTCA)n"
 9890..9923
 repeat_region
 /rpt_family="AT-rich"
 10066..10473
 repeat_region
 /rpt_family="L2"
 10474..10520
 repeat_region
 /rpt_family="CA)n"
 10521..10746
 repeat_region
 /rpt_family="L2"
 12200..12385
 repeat_region
 /rpt_family="MIR"
 12452..12489
 repeat_region
 /rpt_family="CACG)n"
 13410..13773
 repeat_region
 /rpt_family="MIR"
 14076..14317
 repeat_region
 /rpt_family="MIR"
 15470..15592
 repeat_region
 /rpt_family="MIR"
 15599..15967
 repeat_region
 /rpt_family="MIR"
 16227..16617
 misc_feature
 /note="similar to EST AA913068 (NID:93052460) o134e06.s1"
 16517..16693
 repeat_region
 /rpt_family="L1"
 16717..16788
 repeat_region
 /rpt_family="L1"
 16988..17009
 repeat_region
 /rpt_family="AT-rich"
 18169..18509
 repeat_region
 /rpt_family="MER103"
 18594..18644
 repeat_region
 /rpt_family="(TG)n"
 19032..19193
 repeat_region
 /rpt_family="MER1_type"
 19359..19668
 repeat_region
 /rpt_family="L2"
 22454..22498
 repeat_region
 /rpt_family="GA-rich"
 25334..25644
 repeat_region
 /rpt_family="Alu"
 25647..25768
 repeat_region
 /rpt_family="MIR"
 25770..25888
 repeat_region
 /rpt_family="L2"
 25893..26193
 repeat_region
 /rpt_family="Alu"
 25910..26358
 misc_feature
 /note="similar to EST AA578847 (NID:92357031) nk66a09.s1"
 26368..26442
 repeat_region
 /rpt_family="MIR"
 26502..26568
 repeat_region
 /rpt_family="MIR"

Query Match	93.7%	Score 1309,	DB 9,	length 128466;
Best Local Similarity	99.6%	Pred. No. 0;		
Matches 1312; Conservative	0;	Mismatches 5;	Indels 0;	Gaps 0;

LOCUS	DEFINITION	VERSION	KEYWORDS	SOURCE	ORGANISM	REFERENCE	AUTHORS	TITLE	JOURNAL
BC007901	Homo sapiens, clone IMAGE:4133786, mRNA, partial cds.	BC007901.1	GI:14043927	Homo sapiens.	Homo sapiens.	Submitted (11-MAY-2001) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2550, USA	NIH-MGC Project URL: http://mgc.nci.nih.gov	Contact: MGC help desk Email: cgapbs-remail.nih.gov Tissue Procurement: ATCC	CDNA Library Preparation: Rubin Laboratory CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL) DNA Sequencing by: National Institutes of Health Intramural Sequencing Center (NISC), Gaithersburg, Maryland; Web site: http://www.nisc.nih.gov/ Contact: nisc_mgc@nigrl.nih.gov nisc_mgc@nigrl.nih.gov Shenaychenko, Y., Wetherby, K.D., Beckstead, G.G., Binkley, C., Benjamin, B., Blakesley, R.W., Bouffard, G.G., Binkley, C., Brooks, S.,
114515	CACCCCTGTGGGCTGTGAAGCAAGCTCAGGCTAGTCTCCCACTGGGGGCTGGGCCCTTC	114515	CACCCCTGTGGGCTGTGAAGCAAGCTCAGGCTAGTCTCCCACTGGGGGCTGGGCCCTTC	114515	CACCCCTGTGGGCTGTGAAGCAAGCTCAGGCTAGTCTCCCACTGGGGGCTGGGCCCTTC	114515	CACCCCTGTGGGCTGTGAAGCAAGCTCAGGCTAGTCTCCCACTGGGGGCTGGGCCCTTC	114515	CACCCCTGTGGGCTGTGAAGCAAGCTCAGGCTAGTCTCCCACTGGGGGCTGGGCCCTTC
839	CTGGGACGGTGTCCCTGGGGCAGCCCCATCCTCTGTGTTCATATAGTGTGAATCTAGCTAA	839	CTGGGACGGTGTCCCTGGGGCAGCCCCATCCTCTGTGTTCATATAGTGTGAATCTAGCTAA	839	CTGGGACGGTGTCCCTGGGGCAGCCCCATCCTCTGTGTTCATATAGTGTGAATCTAGCTAA	839	CTGGGACGGTGTCCCTGGGGCAGCCCCATCCTCTGTGTTCATATAGTGTGAATCTAGCTAA	839	CTGGGACGGTGTCCCTGGGGCAGCCCCATCCTCTGTGTTCATATAGTGTGAATCTAGCTAA
114575	CTGGGACGGTGTCCCTGGGGCAGCCCCATCCTCTGTGTTCATATAGTGTGAATCTAGCTAA	114575	CTGGGACGGTGTCCCTGGGGCAGCCCCATCCTCTGTGTTCATATAGTGTGAATCTAGCTAA	114575	CTGGGACGGTGTCCCTGGGGCAGCCCCATCCTCTGTGTTCATATAGTGTGAATCTAGCTAA	114575	CTGGGACGGTGTCCCTGGGGCAGCCCCATCCTCTGTGTTCATATAGTGTGAATCTAGCTAA	114575	CTGGGACGGTGTCCCTGGGGCAGCCCCATCCTCTGTGTTCATATAGTGTGAATCTAGCTAA
899	GCCCCCTGTGCTGTCTGTGCACATGCCACAGCAGCCGGTGGGGGCTGGCTGGGGACATTC	899	GCCCCCTGTGCTGTCTGTGCACATGCCACAGCAGCCGGTGGGGGCTGGCTGGGGACATTC	899	GCCCCCTGTGCTGTCTGTGCACATGCCACAGCAGCCGGTGGGGGCTGGCTGGGGACATTC	899	GCCCCCTGTGCTGTCTGTGCACATGCCACAGCAGCCGGTGGGGGCTGGCTGGGGACATTC	899	GCCCCCTGTGCTGTCTGTGCACATGCCACAGCAGCCGGTGGGGGCTGGCTGGGGACATTC
114635	GCCCCCTGTGCTGTCTGTGCACATGCCACAGCAGCCGGTGGGGGCTGGCTGGGGACATTC	114635	GCCCCCTGTGCTGTCTGTGCACATGCCACAGCAGCCGGTGGGGGCTGGCTGGGGACATTC	114635	GCCCCCTGTGCTGTCTGTGCACATGCCACAGCAGCCGGTGGGGGCTGGCTGGGGACATTC	114635	GCCCCCTGTGCTGTCTGTGCACATGCCACAGCAGCCGGTGGGGGCTGGCTGGGGACATTC	114635	GCCCCCTGTGCTGTCTGTGCACATGCCACAGCAGCCGGTGGGGGCTGGCTGGGGACATTC
959	CATCGTGAAGTGTCTCTCAGCTTAGTGTGGAACAGAGACTTGGCGGGGAGATGCTCCAG	959	CATCGTGAAGTGTCTCTCAGCTTAGTGTGGAACAGAGACTTGGCGGGGAGATGCTCCAG	959	CATCGTGAAGTGTCTCTCAGCTTAGTGTGGAACAGAGACTTGGCGGGGAGATGCTCCAG	959	CATCGTGAAGTGTCTCTCAGCTTAGTGTGGAACAGAGACTTGGCGGGGAGATGCTCCAG	959	CATCGTGAAGTGTCTCTCAGCTTAGTGTGGAACAGAGACTTGGCGGGGAGATGCTCCAG
114695	CATCGTGAAGTGTCTCTCAGCTTAGTGTGGAACAGAGACTTGGCGGGGAGATGCTCCAG	114695	CATCGTGAAGTGTCTCTCAGCTTAGTGTGGAACAGAGACTTGGCGGGGAGATGCTCCAG	114695	CATCGTGAAGTGTCTCTCAGCTTAGTGTGGAACAGAGACTTGGCGGGGAGATGCTCCAG	114695	CATCGTGAAGTGTCTCTCAGCTTAGTGTGGAACAGAGACTTGGCGGGGAGATGCTCCAG	114695	CATCGTGAAGTGTCTCTCAGCTTAGTGTGGAACAGAGACTTGGCGGGGAGATGCTCCAG
1019	GATGTGGGTGATTTCTGTACCTGGGGAGGCTATCTTCTGACCTTCCGACAGGGGACATCTCC	1019	GATGTGGGTGATTTCTGTACCTGGGGAGGCTATCTTCTGACCTTCCGACAGGGGACATCTCC	1019	GATGTGGGTGATTTCTGTACCTGGGGAGGCTATCTTCTGACCTTCCGACAGGGGACATCTCC	1019	GATGTGGGTGATTTCTGTACCTGGGGAGGCTATCTTCTGACCTTCCGACAGGGGACATCTCC	1019	GATGTGGGTGATTTCTGTACCTGGGGAGGCTATCTTCTGACCTTCCGACAGGGGACATCTCC
114755	GATGTGGGTGATTTCTGTACCTGGGGAGGCTATCTTCTGACCTTCCGACAGGGGACATCTCC	114755	GATGTGGGTGATTTCTGTACCTGGGGAGGCTATCTTCTGACCTTCCGACAGGGGACATCTCC	114755	GATGTGGGTGATTTCTGTACCTGGGGAGGCTATCTTCTGACCTTCCGACAGGGGACATCTCC	114755	GATGTGGGTGATTTCTGTACCTGGGGAGGCTATCTTCTGACCTTCCGACAGGGGACATCTCC	114755	GATGTGGGTGATTTCTGTACCTGGGGAGGCTATCTTCTGACCTTCCGACAGGGGACATCTCC
1079	AGGCGACGCGCCAGGGGCTCAGGGGGCAGAGGTGCACACTTACGATATAGCCAAAGCTGGGGTC	1079	AGGCGACGCGCCAGGGGCTCAGGGGGCAGAGGTGCACACTTACGATATAGCCAAAGCTGGGGTC	1079	AGGCGACGCGCCAGGGGCTCAGGGGGCAGAGGTGCACACTTACGATATAGCCAAAGCTGGGGTC	1079	AGGCGACGCGCCAGGGGCTCAGGGGGCAGAGGTGCACACTTACGATATAGCCAAAGCTGGGGTC	1079	AGGCGACGCGCCAGGGGCTCAGGGGGCAGAGGTGCACACTTACGATATAGCCAAAGCTGGGGTC
114815	AGGCGACGCGCCAGGGGCTCAGGGGGCAGAGGTGCACACTTACGATATAGCCAAAGCTGGGGTC	114815	AGGCGACGCGCCAGGGGCTCAGGGGGCAGAGGTGCACACTTACGATATAGCCAAAGCTGGGGTC	114815	AGGCGACGCGCCAGGGGCTCAGGGGGCAGAGGTGCACACTTACGATATAGCCAAAGCTGGGGTC	114815	AGGCGACGCGCCAGGGGCTCAGGGGGCAGAGGTGCACACTTACGATATAGCCAAAGCTGGGGTC	114815	AGGCGACGCGCCAGGGGCTCAGGGGGCAGAGGTGCACACTTACGATATAGCCAAAGCTGGGGTC
1139	AGGAGACAGGTGTGTGTGACCCAGAGACTTGGGGGCGGGGAGCTTCTCTGC	1139	AGGAGACAGGTGTGTGTGACCCAGAGACTTGGGGGCGGGGAGCTTCTCTGC	1139	AGGAGACAGGTGTGTGTGACCCAGAGACTTGGGGGCGGGGAGCTTCTCTGC	1139	AGGAGACAGGTGTGTGTGACCCAGAGACTTGGGGGCGGGGAGCTTCTCTGC	1139	AGGAGACAGGTGTGTGTGACCCAGAGACTTGGGGGCGGGGAGCTTCTCTGC
114875	AGGAGACAGGTGTGTGTGACCCAGAGACTTGGGGGCGGGGAGCTTCTCTGC	114875	AGGAGACAGGTGTGTGTGACCCAGAGACTTGGGGGCGGGGAGCTTCTCTGC	114875	AGGAGACAGGTGTGTGTGACCCAGAGACTTGGGGGCGGGGAGCTTCTCTGC	114875	AGGAGACAGGTGTGTGTGACCCAGAGACTTGGGGGCGGGGAGCTTCTCTGC	114875	AGGAGACAGGTGTGTGTGACCCAGAGACTTGGGGGCGGGGAGCTTCTCTGC
1199	CTCATTTGCTTTCATGAAAGACCTTCAAGACGCCAAACCAAGGCTTTTCCCTCTCGA	1199	CTCATTTGCTTTCATGAAAGACCTTCAAGACGCCAAACCAAGGCTTTTCCCTCTCGA	1199	CTCATTTGCTTTCATGAAAGACCTTCAAGACGCCAAACCAAGGCTTTTCCCTCTCGA	1199	CTCATTTGCTTTCATGAAAGACCTTCAAGACGCCAAACCAAGGCTTTTCCCTCTCGA	1199	CTCATTTGCTTTCATGAAAGACCTTCAAGACGCCAAACCAAGGCTTTTCCCTCTCGA
114935	CTCATTTGCTTTCATGAAAGACCTTCAAGACGCCAAACCAAG								





CC medical conditions e.g. by protein or gene therapy, especially
CC disorders related to the excretory system. The genes are isolated
CC from a range of human tissues disclosed in the specification. The
CC nucleic acids, proteins, antibodies and (ant)agonists are useful in
CC the diagnosis, treatment and prevention of: (a) cancer, e.g. breast
CC and ovarian cancer and other cancers of the adrenal gland, bone, bone
CC marrow, breast, gastrointestinal tract, liver, lung, or urogenital;
CC (b) immune disorders e.g. Addison's disease, allergies, autoimmune
CC haemolytic anaemia, autoimmune thyroiditis, diabetes mellitus, Crohn's
CC disease, multiple sclerosis, rheumatoid arthritis and ulcerative
CC colitis; (c) cardiovascular disorders such as myocardial ischaemia;
CC epilepsy; and (f) infectious diseases e.g. cerebral anoxia and
CC and parasitic infections.
CC Note: The sequence data for this patent did not form part of the
CC printed specification, but was obtained in electronic format directly
CC from W10 at ftp.wipo.int/pub/published_pcc_sequences.
XX

Sequence 9839 BP; 2292 A; 2702 C; 2654 G; 2191 T; 0 other;

Query Match 92.0%; Score 1284.8; DB 22; Length 9839;
Best Local Similarity 99.7%; Pred. No. 1.1e-289;

Matches 1308; Conservative 0; Mismatches 2; Indels 2; Gaps 2;

59 ACCGAGTACCAAGTATATCCAGACAGACCATCCCGAGAGAGACTACCGCTGGGCA 118
8530 AAGAGTACCAAGTATATCCAGACAGACCATCCCGAGAGAGACTACCGCTGGGCA 8589
119 TCTTACCAAGTATATCCAGACAGACCATCCCGAGAGAGACTACCGCTGGGCA 178
8530 TCTTACCAAGTATATCCAGACAGACCATCCCGAGAGAGACTACCGCTGGGCA 8649
179 TGTGAGAGCCAGTATATCCAGACAGACCATCCCGAGAGAGACTACCGCTGGGCA 238
8650 TGTGAGAGCCAGTATATCCAGACAGACCATCCCGAGAGAGACTACCGCTGGGCA 8709
239 GAGCCAGTATATCCAGACAGACCATCCCGAGAGAGACTACCGCTGGGCA 298
8710 GAGCCAGTATATCCAGACAGACCATCCCGAGAGAGACTACCGCTGGGCA 8769
299 AGTATATCCAGTATATCCAGACAGACCATCCCGAGAGAGACTACCGCTGGGCA 358
8770 AGTATATCCAGTATATCCAGACAGACCATCCCGAGAGAGACTACCGCTGGGCA 8829
359 CTTCAAGGAGTATATCCAGACAGACCATCCCGAGAGAGACTACCGCTGGGCA 418
8830 CTTCAAGGAGTATATCCAGACAGACCATCCCGAGAGAGACTACCGCTGGGCA 8889
419 TGTATCTCTACCTCTGAGGAGTATATCCAGACAGACCATCCCGAGAGAGACTACCGCTGGGCA 478
8890 TGTATCTCTACCTCTGAGGAGTATATCCAGACAGACCATCCCGAGAGAGACTACCGCTGGGCA 8949
479 CCAAGTATATCCAGTATATCCAGACAGACCATCCCGAGAGAGACTACCGCTGGGCA 538
8950 CCAAGTATATCCAGTATATCCAGACAGACCATCCCGAGAGAGACTACCGCTGGGCA 9008
539 AGGCTGAGTATATCCAGTATATCCAGACAGACCATCCCGAGAGAGACTACCGCTGGGCA 598
9009 AGGCTGAGTATATCCAGTATATCCAGACAGACCATCCCGAGAGAGACTACCGCTGGGCA 9068
599 TTGCACTGAGTATATCCAGTATATCCAGACAGACCATCCCGAGAGAGACTACCGCTGGGCA 658
9069 TTGCACTGAGTATATCCAGTATATCCAGACAGACCATCCCGAGAGAGACTACCGCTGGGCA 9128
659 CAGACAGTATATCCAGTATATCCAGACAGACCATCCCGAGAGAGACTACCGCTGGGCA 718
9129 CAGACAGTATATCCAGTATATCCAGACAGACCATCCCGAGAGAGACTACCGCTGGGCA 9188
719 AGCTAGTATATCCAGTATATCCAGACAGACCATCCCGAGAGAGACTACCGCTGGGCA 778
9189 AGCTAGTATATCCAGTATATCCAGACAGACCATCCCGAGAGAGACTACCGCTGGGCA 9248
779 CAACTGAGTATATCCAGTATATCCAGACAGACCATCCCGAGAGAGACTACCGCTGGGCA 838

DB 9249 CAACTGAGTATATCCAGTATATCCAGACAGACCATCCCGAGAGAGACTACCGCTGGGCA 9307
9249 CAACTGAGTATATCCAGTATATCCAGACAGACCATCCCGAGAGAGACTACCGCTGGGCA 9367
839 CTGGAGAGTATATCCAGTATATCCAGACAGACCATCCCGAGAGAGACTACCGCTGGGCA 898
9308 CTGGAGAGTATATCCAGTATATCCAGACAGACCATCCCGAGAGAGACTACCGCTGGGCA 9367
899 GCGCTGAGTATATCCAGTATATCCAGACAGACCATCCCGAGAGAGACTACCGCTGGGCA 958
9368 GCGCTGAGTATATCCAGTATATCCAGACAGACCATCCCGAGAGAGACTACCGCTGGGCA 9427
959 CATCTGAGTATATCCAGTATATCCAGACAGACCATCCCGAGAGAGACTACCGCTGGGCA 1018
9428 CATCTGAGTATATCCAGTATATCCAGACAGACCATCCCGAGAGAGACTACCGCTGGGCA 9487
1019 GATGAGTATATCCAGTATATCCAGACAGACCATCCCGAGAGAGACTACCGCTGGGCA 1078
9488 GATGAGTATATCCAGTATATCCAGACAGACCATCCCGAGAGAGACTACCGCTGGGCA 9547
1079 AGGCAAGTATATCCAGTATATCCAGACAGACCATCCCGAGAGAGACTACCGCTGGGCA 1138
9548 AGGCAAGTATATCCAGTATATCCAGACAGACCATCCCGAGAGAGACTACCGCTGGGCA 9607
1139 AGGCAAGTATATCCAGTATATCCAGACAGACCATCCCGAGAGAGACTACCGCTGGGCA 1198
9608 AGGCAAGTATATCCAGTATATCCAGACAGACCATCCCGAGAGAGACTACCGCTGGGCA 9667
1199 CTGATGAGTATATCCAGTATATCCAGACAGACCATCCCGAGAGAGACTACCGCTGGGCA 1258
9668 CTGATGAGTATATCCAGTATATCCAGACAGACCATCCCGAGAGAGACTACCGCTGGGCA 9727
1259 GTTGAATATATCCAGTATATCCAGACAGACCATCCCGAGAGAGACTACCGCTGGGCA 1318
9728 GTTGAATATATCCAGTATATCCAGACAGACCATCCCGAGAGAGACTACCGCTGGGCA 9787
1319 ATAAATATATCCAGTATATCCAGACAGACCATCCCGAGAGAGACTACCGCTGGGCA 1370
9788 ATAAATATATCCAGTATATCCAGACAGACCATCCCGAGAGAGACTACCGCTGGGCA 9839

RESULT 2

AA163402 standard DNA; 9839 BP.

AA163402;

22-OCT-2001 (first entry).

Human kidney related polynucleotide SEQ ID NO 717.

Human; kidney antigen; immunosuppressive; antiarthritic; antirheumatic;
antiproliferative; cytotoxic; cardiant; vasotropic; cerebroprotective;
neurotropic; neuroprotective; antibacterial; vitucide; fungicide;
ophthalmitic; antiallergic; hepatotropic; antidiabetic;
antiinflammatory; antitumor; vulnery; anticonvulsant; antiparasitic;
gene therapy; cancer; immune disorder; cardiovascular disorder;
neurological disease; infection; ds.

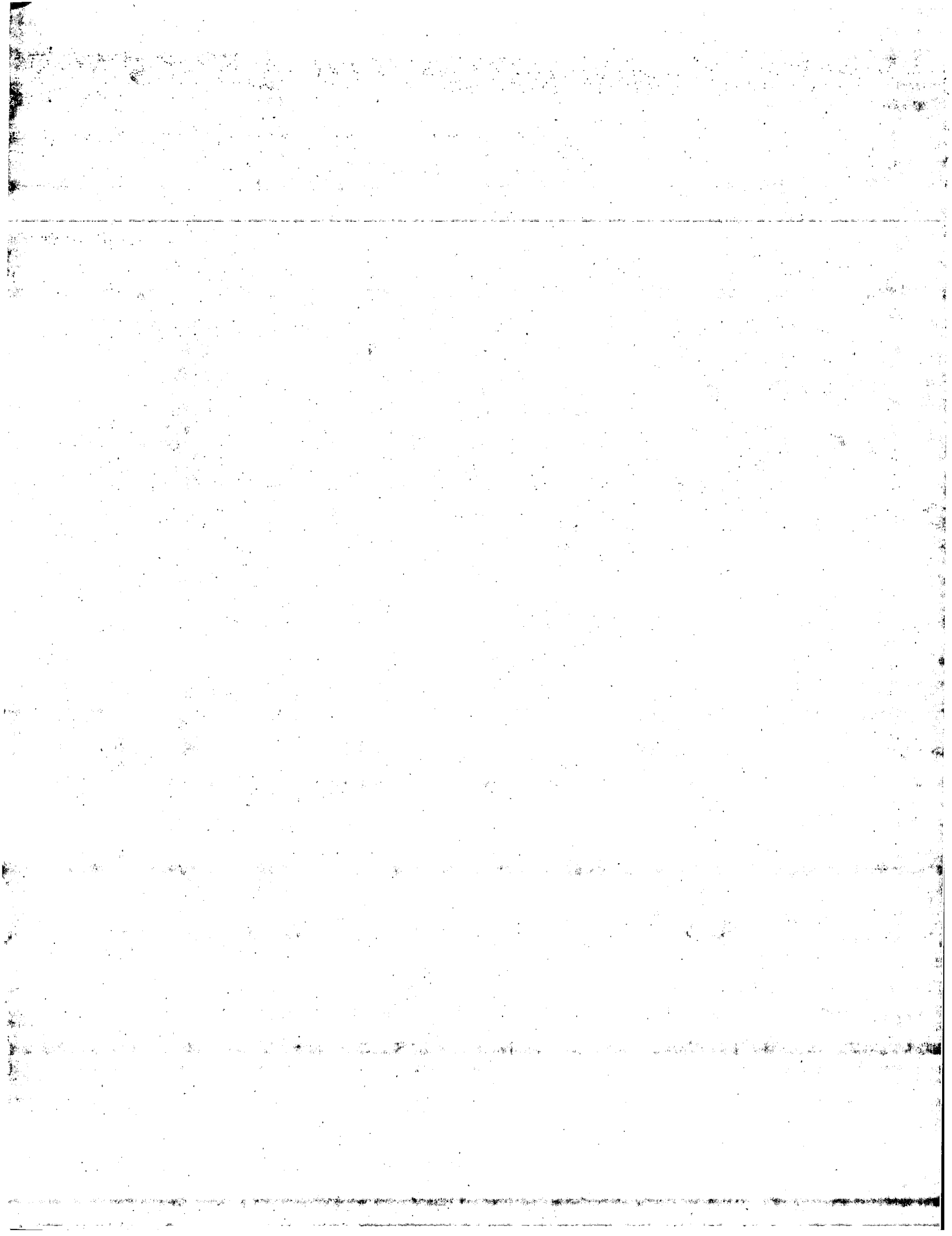
Homo sapiens.

WO200155323-A2.

02-AUG-2001.

17-JAN-2001; 2001MO-US01343.

31-JAN-2000; 2000US-0179065.
PR 04-FEB-2000; 2000US-0180628.
PR 24-FEB-2000; 2000US-0184664.
PR 02-MAR-2000; 2000US-0186350.
PR 16-MAR-2000; 2000US-0189874.
PR 17-MAR-2000; 2000US-0190076.
PR 18-APR-2000; 2000US-0198123.



PR	19-MAY-2000,	2000US-0205515
PR	07-JUN-2000,	2000US-0209677
PR	28-JUN-2000,	2000US-0214886
PR	30-JUN-2000,	2000US-0215135
PR	07-JUL-2000,	2000US-0216647
PR	07-JUL-2000,	2000US-0216880
PR	11-JUL-2000,	2000US-0217486
PR	11-JUL-2000,	2000US-0217497
PR	26-JUL-2000,	2000US-0218290
PR	26-JUL-2000,	2000US-0220963
PR	26-JUL-2000,	2000US-0220964
PR	14-AUG-2000,	2000US-0224518
PR	14-AUG-2000,	2000US-0224519
PR	14-AUG-2000,	2000US-0225213
PR	14-AUG-2000,	2000US-0225214
PR	14-AUG-2000,	2000US-0225266
PR	14-AUG-2000,	2000US-0225267
PR	14-AUG-2000,	2000US-0225268
PR	14-AUG-2000,	2000US-0225270
PR	14-AUG-2000,	2000US-0225447
PR	14-AUG-2000,	2000US-0225757
PR	14-AUG-2000,	2000US-0225758
PR	18-AUG-2000,	2000US-0225755
PR	18-AUG-2000,	2000US-0225756
PR	22-AUG-2000,	2000US-0226279
PR	22-AUG-2000,	2000US-0226281
PR	22-AUG-2000,	2000US-0226868
PR	23-AUG-2000,	2000US-0227182
PR	30-AUG-2000,	2000US-0227009
PR	01-SEP-2000,	2000US-0228924
PR	01-SEP-2000,	2000US-0229287
PR	01-SEP-2000,	2000US-0229343
PR	01-SEP-2000,	2000US-0229344
PR	01-SEP-2000,	2000US-0229345
PR	05-SEP-2000,	2000US-0229509
PR	05-SEP-2000,	2000US-0229513
PR	06-SEP-2000,	2000US-0230437
PR	06-SEP-2000,	2000US-0230438
PR	08-SEP-2000,	2000US-0231242
PR	08-SEP-2000,	2000US-0231243
PR	08-SEP-2000,	2000US-0231244
PR	08-SEP-2000,	2000US-0231413
PR	08-SEP-2000,	2000US-0231414
PR	08-SEP-2000,	2000US-0232080
PR	12-SEP-2000,	2000US-0232081
PR	12-SEP-2000,	2000US-0232357
PR	14-SEP-2000,	2000US-0232358
PR	14-SEP-2000,	2000US-0232359
PR	14-SEP-2000,	2000US-0232400
PR	14-SEP-2000,	2000US-0232401
PR	14-SEP-2000,	2000US-0233063
PR	14-SEP-2000,	2000US-0233064
PR	21-SEP-2000,	2000US-0233065
PR	21-SEP-2000,	2000US-0234223
PR	21-SEP-2000,	2000US-0234274
PR	25-SEP-2000,	2000US-0234997
PR	25-SEP-2000,	2000US-0234998
PR	26-SEP-2000,	2000US-0235498
PR	27-SEP-2000,	2000US-0235494
PR	27-SEP-2000,	2000US-0235834
PR	29-SEP-2000,	2000US-0236337
PR	29-SEP-2000,	2000US-0236337
PR	29-SEP-2000,	2000US-0236338
PR	29-SEP-2000,	2000US-0236369
PR	02-OCT-2000,	2000US-0236870
PR	02-OCT-2000,	2000US-0236802
PR	02-OCT-2000,	2000US-0237037
PR	02-OCT-2000,	2000US-0237038
PR	02-OCT-2000,	2000US-0237039
PR	13-OCT-2000,	2000US-0237040
PR	13-OCT-2000,	2000US-0239935
PR	13-OCT-2000,	2000US-0239937
PR	30-OCT-2000,	2000US-0240960

PR 20-OCT-2000; 2000US-0241221.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 08-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249246.
PR 17-NOV-2000; 2000US-0249247.
PR 17-NOV-2000; 2000US-0249248.
PR 17-NOV-2000; 2000US-0249249.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249299.
PR 17-NOV-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250160.
PR 01-DEC-2000; 2000US-0250391.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.
PR 05-DEC-2000; 2000US-0256719.
PR 06-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251868.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251989.
PR 08-DEC-2000; 2000US-0251990.
PR 11-DEC-2000; 2000US-0254097.
PR 05-JAN-2001; 2001US-0259676.

XX
XX PA (HUMA-) HUMAN GENOME SCI INC.
XX PI Rosen CA, Barash SC, Ruben SW;
XX DR WPI; 2001-488784/53.
XX
CC The invention relates to novel kidney related polynucleotides
CC (AAI62971-AAI6793) and the encoded polypeptides (AAW42417-AAW42691)
CC collectively known as kidney antigens and the use of such kidney antigens
CC for detecting disorders of the kidney, especially kidney cancer and
CC kidney cancer metastases. The polymy

useful for preventing, treating or ameliorating medical conditions e.g. by protein or gene therapy. The genes are isolated from a range of human tissues disclosed in the specification. The nucleic acids, proteins, antibodies and (ant)agonists are useful in the diagnosis, treatment and prevention of: (a) cancer, e.g. breast and ovarian cancer, and other cancers of the adrenal gland, bone, bone marrow, breast, gastrointestinal tract, liver, lung, or urogenital; (b) immune disorders e.g. Addison's disease, allergies, autoimmune haemolytic anaemia, autoimmune thyroiditis, diabetes mellitus, Crohn's disease, multiple sclerosis, rheumatoid arthritis and ulcerative colitis; (c) cardiovascular disorders such as myocardial ischaemia; (d) wound healing; (e) neurological diseases e.g. cerebral anoxia and epilepsy; and (f) infectious diseases such as viral, bacterial, fungal and parasitic infections.

Note: The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from WIPO at ftp.wipo.int/pub/published_pct_sequences.

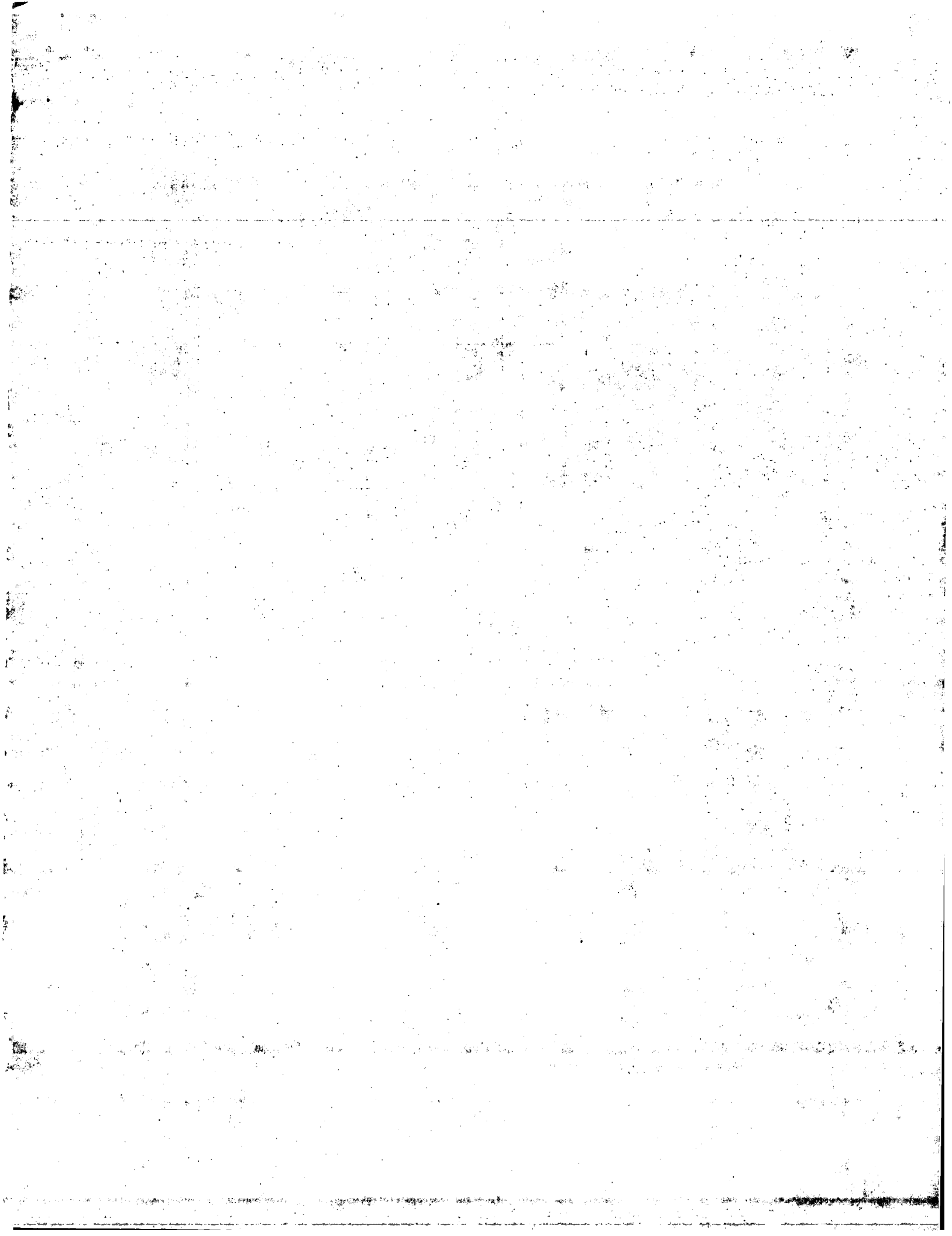
Sequence 9839 BP, 2292 A; 2702 C; 2654 G; 2191 T; 0 other;

Query Match 92.0% Score 1284.8; DB 22; Length 9839;
Best Local Similarity 99.7%; Pred. No. 1.1e-289;
Matches 1308; Conservative 0; Mismatches 2; Indels 2; Gaps 2;

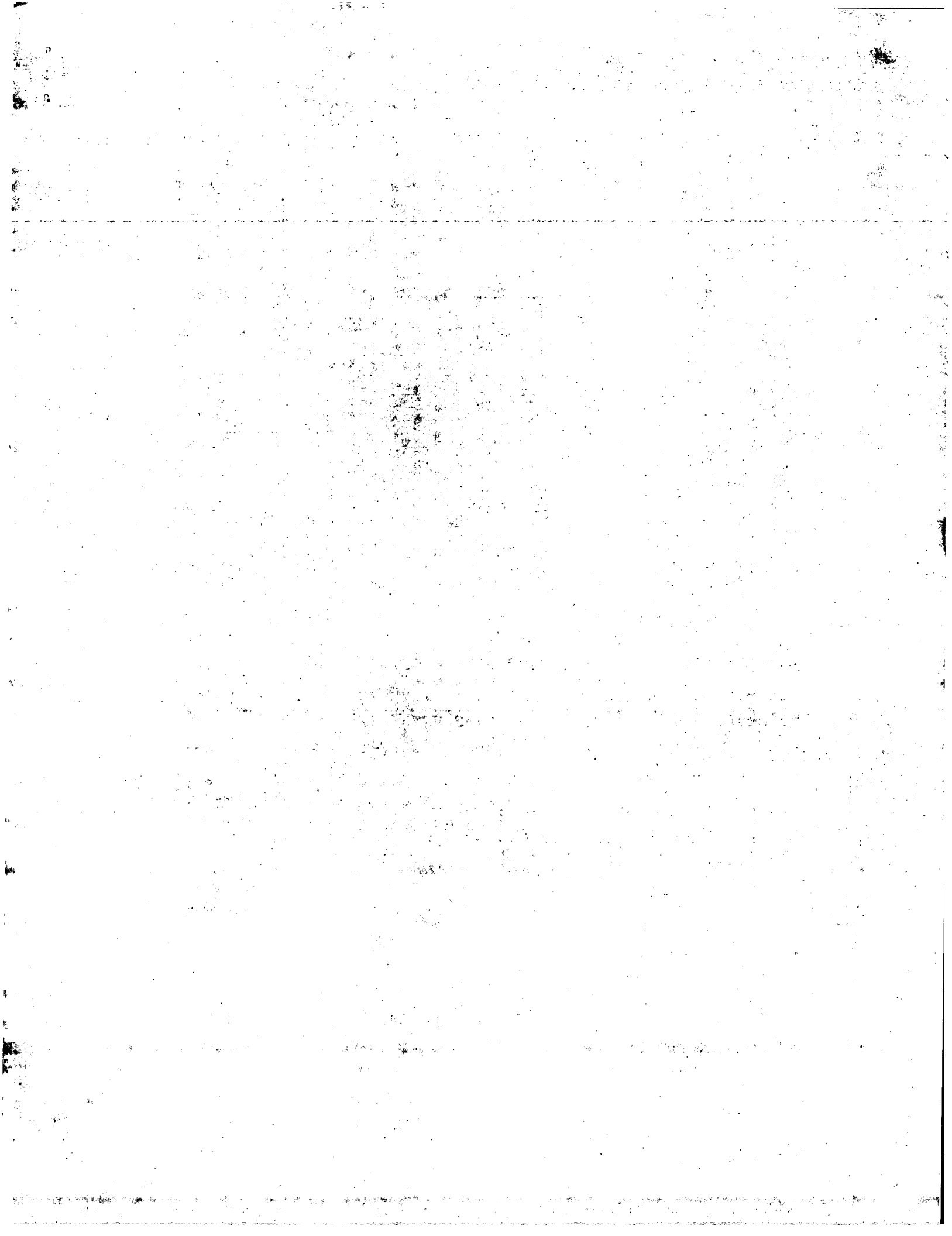
QY 59 ACCGAGTACAGTGTATCCAGACAGACATCCCGGAGAACTACCGCTGCGCA 118
DB 8530 ACAGAGTACAGTGTATCCAGACAGACATCCCGGAGAACTACCGCTGCGCA 8589
QY 119 TCTTACACACCGGAGAGTGTCTCTTCACTGTTCACCTGCTGAGCTGTGATGTC 178
DB 8590 TCTTACACACCGGAGAGTGTCTCTTCACTGTTCACCTGCTGAGCTGTGATGTC 8649
QY 179 TGTGAGAGCAATGCGGAGTGTCTCTTCACTGTTCACCTGCTGAGCTGTGATGTC 238
DB 8650 TGTGAGAGCAATGCGGAGTGTCTCTTCACTGTTCACCTGCTGAGCTGTGATGTC 8709
QY 239 GAGCAGTGTGAGAGAGCTCTTCAAGGAGATGAGAGAGAGAGAGAGAGAGAGAT 298
DB 8710 GAGCAGTGTGAGAGAGAGCTCTTCAAGGAGATGAGAGAGAGAGAGAGAGAGAT 8769
QY 299 AGTGTATCCCATGAGAGTGTCTCTTCAAGGAGATGAGAGAGAGAGAGAGAT 358
DB 8770 AGTGTATCCCATGAGAGTGTCTCTTCAAGGAGATGAGAGAGAGAGAGAGAT 8829
QY 359 CTTCAGAGGAGTCAATTAAGGAGAGATGCTTCTTCAAGGAGAGAGAGAGAGAT 418
DB 8830 CTTCAGAGGAGTCAATTAAGGAGAGATGCTTCTTCAAGGAGAGAGAGAGAT 8889
QY 419 TGTATCTCTACCTGTGCGGAGTGTCTTCTTCAAGGAGAGAGAGAGAT 478
DB 8890 TGTATCTCTACCTGTGCGGAGTGTCTTCTTCAAGGAGAGAGAGAGAT 8949
QY 479 CCAAGTGTCTCTGATCCCAAGAGACATATATGAGAGAGAGAGAGAGAT 538
DB 8950 CCAAGTGTCTCTGATCCCAAGAGACATATATGAGAGAGAGAGAGAGAT 9008
QY 539 AGGCTCTGCTGACAGAGTGTCTCTTCAAGGAGAGAGAGAGAGAT 598
DB 9009 AGGCTCTGCTGACAGAGTGTCTCTTCAAGGAGAGAGAGAGAGAT 9068
QY 599 TTGCACTGACAGAGTGTCTCTTCAAGGAGAGAGAGAGAGAT 658
DB 9069 TTGCACTGACAGAGTGTCTCTTCAAGGAGAGAGAGAGAGAT 9128
QY 659 CAGACAGATGTGACAGAGACAACTGTCAATATGCAATATGATGATGAT 718
DB 9129 CAGACAGATGTGACAGAGACAACTGTCAATATGCAATATGATGATGAT 9188
QY 719 AGCTTACGATGAGAGTGTCTCTTCAAGGAGAGAGAGAGAT 778
DB 9189 AGCTTACGATGAGAGTGTCTCTTCAAGGAGAGAGAGAGAT 9248
QY 779 CAACCTGTGAGCTGTATGAGAGAGTGTCTCTTCAAGGAGAGAGAGAT 838

DB 9249 CAACCTGTGAGCTGTATGAGAGAGTGTCTCTTCAAGGAGAGAGAGAT 9307
QY 839 CTGGAGAGCTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAT 898
DB 9308 CTGGAGAGCTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAT 9367
QY 899 GCCCTGTGCTGTGCTGTGACATGACAGAGAGAGAGAGAGAGAGAGAT 958
DB 9368 GCCCTGTGCTGTGCTGTGACATGACAGAGAGAGAGAGAGAGAGAGAT 9427
QY 959 CATGTGAGTGTCTGTCTGAGCTTGTGAGAGAGAGAGAGAGAGAGAT 1018
DB 9428 CATGTGAGTGTCTGTCTGAGCTTGTGAGAGAGAGAGAGAGAGAT 9487
QY 1019 GATGTGAGTGTCTGTCTGAGCTTGTGAGAGAGAGAGAGAGAGAGAT 1078
DB 9488 GATGTGAGTGTCTGTCTGAGCTTGTGAGAGAGAGAGAGAGAGAT 9547
QY 1079 AGGCAAGCCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAT 1138
DB 9548 AGGCAAGCCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAT 9607
QY 1139 AGGAGAGAGTGTGTGTTGAGAGAGAGAGAGAGAGAGAGAGAT 1198
DB 9608 AGGAGAGAGTGTGTGTTGAGAGAGAGAGAGAGAGAGAGAGAT 9667
QY 1199 CTGATTTGCTTCAAGAGAGAGAGAGAGAGAGAGAGAGAGAT 1258
DB 9668 CTGATTTGCTTCAAGAGAGAGAGAGAGAGAGAGAGAGAT 9727
QY 1259 GTTGAATATCAGAGATCTTGTGATCTTGTGATTAATTTTGTATAAA 1318
DB 9728 GTTGAATATCAGAGATCTTGTGATCTTGTGATTAATTTTGTATAAA 9787
QY 1319 ATAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1370
DB 9788 ATAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 9839

RESULT 3
AAH98172
ID AAH98172 standard; cDNA, 1846 BP.
XX
AC AAH98172;
XX
DT 12-OCT-2001 (first entry)
XX
DE Human EST-derived coding sequence SEQ ID NO: 29.
XX
KW Human: sheep; pig; cow; fruit fly; yeast; hamster; macaque; horse;
KW tomato; monkey; dog; sea urchin; expressed sequence tag; EST;
KW diagnostics; forensic test; gene mapping; genetic disorder;
XX
OS Homo sapiens.
XX
PN WO200154477-A2.
XX
PD 02-AUG-2001.
XX
PF 25-JAN-2001; 2001MO-US02687.
XX
PR 25-JAN-2000; 2000US-0491404.
PR 17-JUL-2000; 2000US-0617746.
PR 03-AUG-2000; 2000US-0631451.
PR 15-SEP-2000; 2000US-0663870.
XX
PA (HYSB-) HYSBQ INC.
XX
PI Tang YF, Liu C, Zhou P, Qian XB, Wang Z, Chen R, Aundt V,
Cao Y, Dymnac RA, Zhang J, Wehrman T;



PR	04-FEB-2000	2000US-0180628
PR	24-FEB-2000	2000US-0184664
PR	02-MAR-2000	2000US-0185350
PR	15-MAR-2000	2000US-0189874
PR	17-MAR-2000	2000US-0190076
PR	18-APR-2000	2000US-0198123
PR	19-MAY-2000	2000US-0205515
PR	07-JUN-2000	2000US-0209467
PR	28-JUN-2000	2000US-0214866
PR	30-JUN-2000	2000US-0215155
PR	07-JUL-2000	2000US-0216647
PR	07-JUL-2000	2000US-0216887
PR	11-JUL-2000	2000US-0217496
PR	14-JUL-2000	2000US-0218220
PR	26-JUL-2000	2000US-0220963
PR	26-JUL-2000	2000US-0220964
PR	14-AUG-2000	2000US-0224518
PR	14-AUG-2000	2000US-0224519
PR	14-AUG-2000	2000US-0225521
PR	14-AUG-2000	2000US-0225524
PR	14-AUG-2000	2000US-0225526
PR	14-AUG-2000	2000US-0225567
PR	14-AUG-2000	2000US-0225568
PR	14-AUG-2000	2000US-0225570
PR	14-AUG-2000	2000US-0225547
PR	14-AUG-2000	2000US-0225575
PR	14-AUG-2000	2000US-0225758
PR	14-AUG-2000	2000US-0225759
PR	18-AUG-2000	2000US-0226579
PR	22-AUG-2000	2000US-0226681
PR	22-AUG-2000	2000US-0226686
PR	22-AUG-2000	2000US-0226688
PR	23-AUG-2000	2000US-0227018
PR	30-ANG-2000	2000US-0228824
PR	01-SEP-2000	2000US-0229287
PR	01-SEP-2000	2000US-0229343
PR	01-SEP-2000	2000US-0229344
PR	01-SEP-2000	2000US-0229345
PR	05-SEP-2000	2000US-0229509
PR	05-SEP-2000	2000US-0229517
PR	06-SEP-2000	2000US-0230438
PR	08-SEP-2000	2000US-0231342
PR	08-SEP-2000	2000US-0231343
PR	08-SEP-2000	2000US-0231343
PR	08-SEP-2000	2000US-0231343
PR	08-SEP-2000	2000US-0231413
PR	08-SEP-2000	2000US-0231414
PR	08-SEP-2000	2000US-0231415
PR	08-SEP-2000	2000US-0231416
PR	08-SEP-2000	2000US-0231417
PR	08-SEP-2000	2000US-0231418
PR	08-SEP-2000	2000US-0231419
PR	08-SEP-2000	2000US-0231420
PR	08-SEP-2000	2000US-0231421
PR	08-SEP-2000	2000US-0231422
PR	08-SEP-2000	2000US-0231423
PR	08-SEP-2000	2000US-0231424
PR	08-SEP-2000	2000US-0231425
PR	08-SEP-2000	2000US-0231426
PR	08-SEP-2000	2000US-0231427
PR	08-SEP-2000	2000US-0231428
PR	08-SEP-2000	2000US-0231429
PR	08-SEP-2000	2000US-0231430
PR	08-SEP-2000	2000US-0231431
PR	08-SEP-2000	2000US-0231432
PR	08-SEP-2000	2000US-0231433
PR	08-SEP-2000	2000US-0231434
PR	08-SEP-2000	2000US-0231435
PR	08-SEP-2000	2000US-0231436
PR	08-SEP-2000	2000US-0231437
PR	08-SEP-2000	2000US-0231438
PR	08-SEP-2000	2000US-0231439
PR	08-SEP-2000	2000US-0231440
PR	08-SEP-2000	2000US-0231441
PR	08-SEP-2000	2000US-0231442
PR	08-SEP-2000	2000US-0231443
PR	08-SEP-2000	2000US-0231444
PR	08-SEP-2000	2000US-0231445
PR	08-SEP-2000	2000US-0231446
PR	08-SEP-2000	2000US-0231447
PR	08-SEP-2000	2000US-0231448
PR	08-SEP-2000	2000US-0231449
PR	08-SEP-2000	2000US-0231450
PR	08-SEP-2000	2000US-0231451
PR	08-SEP-2000	2000US-0231452
PR	08-SEP-2000	2000US-0231453
PR	08-SEP-2000	2000US-0231454
PR	08-SEP-2000	2000US-0231455
PR	08-SEP-2000	2000US-0231456
PR	08-SEP-2000	2000US-0231457
PR	08-SEP-2000	2000US-0231458
PR	08-SEP-2000	2000US-0231459
PR	08-SEP-2000	2000US-0231460
PR	08-SEP-2000	2000US-0231461
PR	08-SEP-2000	2000US-0231462



02-OCT-2000; 2000US-0237038.
 02-OCT-2000; 2000US-0237039.
 13-OCT-2000; 2000US-0237040.
 13-OCT-2000; 2000US-0239335.
 13-OCT-2000; 2000US-0239337.
 20-OCT-2000; 2000US-0240960.
 20-OCT-2000; 2000US-0241221.
 20-OCT-2000; 2000US-0241785.
 20-OCT-2000; 2000US-0241866.
 20-OCT-2000; 2000US-0241877.
 20-OCT-2000; 2000US-0241808.
 20-OCT-2000; 2000US-0241809.
 20-OCT-2000; 2000US-0241826.
 01-NOV-2000; 2000US-0244117.
 08-NOV-2000; 2000US-0244674.
 08-NOV-2000; 2000US-0246475.
 08-NOV-2000; 2000US-0246476.
 08-NOV-2000; 2000US-0246477.
 08-NOV-2000; 2000US-0246478.
 08-NOV-2000; 2000US-0246523.
 08-NOV-2000; 2000US-0246524.
 08-NOV-2000; 2000US-0246525.
 08-NOV-2000; 2000US-0246526.
 08-NOV-2000; 2000US-0246527.
 08-NOV-2000; 2000US-0246528.
 08-NOV-2000; 2000US-0246532.
 08-NOV-2000; 2000US-0246609.
 08-NOV-2000; 2000US-0246610.
 08-NOV-2000; 2000US-0246611.
 08-NOV-2000; 2000US-0246613.
 17-NOV-2000; 2000US-0249207.
 17-NOV-2000; 2000US-0249208.
 17-NOV-2000; 2000US-0249209.
 17-NOV-2000; 2000US-0249210.
 17-NOV-2000; 2000US-0249211.
 17-NOV-2000; 2000US-0249212.
 17-NOV-2000; 2000US-0249213.
 17-NOV-2000; 2000US-0249214.
 17-NOV-2000; 2000US-0249215.
 17-NOV-2000; 2000US-0249216.
 17-NOV-2000; 2000US-0249217.
 17-NOV-2000; 2000US-0249218.
 17-NOV-2000; 2000US-0249244.
 17-NOV-2000; 2000US-0249245.
 17-NOV-2000; 2000US-0249264.
 17-NOV-2000; 2000US-0249265.
 17-NOV-2000; 2000US-0249297.
 17-NOV-2000; 2000US-0249299.
 17-NOV-2000; 2000US-0249300.
 01-DEC-2000; 2000US-0250160.
 01-DEC-2000; 2000US-0250391.
 05-DEC-2000; 2000US-0251030.
 05-DEC-2000; 2000US-0251988.
 05-DEC-2000; 2000US-0256179.
 06-DEC-2000; 2000US-0251479.
 08-DEC-2000; 2000US-0251856.
 08-DEC-2000; 2000US-0251868.
 08-DEC-2000; 2000US-0251869.
 08-DEC-2000; 2000US-0251989.
 08-DEC-2000; 2000US-0251990.
 11-DEC-2000; 2000US-0254097.
 05-JAN-2001; 2001US-0259678.
 (HUMA-) HUMAN GENOME SCI INC.
 Rosen CA, Barash SC, Ruben SM;
 WPI; 2001-488784/53.
 New isolated nucleic acids and polypeptides, useful for diagnosing,
 treating and/or preventing human diseases and disorders -
 Disclosure; SEQ ID NO 717; 564pp + Sequence Listing; English.

XX The invention relates to novel kidney related polynucleotides
 CC (AA162971-AA163793) and the encoded polypeptides (AA162417-AA162691)
 CC collectively known as kidney antigens and the use of such kidney antigens
 CC for detecting disorders of the kidney, especially kidney cancer and
 CC kidney cancer metastases. The polynucleotides and proteins are also
 CC useful for preventing, treating or ameliorating medical conditions
 CC e.g. by protein or gene therapy. The genes are isolated from a range
 CC of human tissues disclosed in the specification. The nucleic acids
 CC proteins, antibodies and (ant)agonists are useful in the diagnosis,
 CC treatment and prevention of: (a) cancer, e.g. breast and ovarian cancer,
 CC and other cancers of the adrenal gland, bone, bone marrow, breast,
 CC gastrointestinal tract, liver, lung, or urogenital; (b) immune disorders
 CC e.g. Addison's disease, allergies, autoimmune haemolytic anaemia,
 CC autoimmune thyroiditis, diabetes mellitus, Crohn's disease, multiple
 CC sclerosis, rheumatoid arthritis and ulcerative colitis;
 CC (c) cardiovascular disorders such as myocardial ischaemias; (d) wound
 CC healing; (e) neurological diseases e.g. cerebral anoxia and epilepsy;
 CC and (f) infectious diseases such as viral, bacterial, fungal and
 CC parasitic infections.
 CC Note: The sequence data for this patent did not form part of the
 CC printed specification, but was obtained in electronic format directly
 CC from WIPO at ftp.wipo.int/pub/published_pat_sequences.
 XX
 SQ Sequence 9839 BP; 2292 A; 2702 C; 2654 G; 2191 T; 0 other;
 Query Match 72.7%; Score 1016; DB 22; Length 9839;
 Best Local Similarity 99.8%; Pred. No. 0;
 Matches 1106; Conservative 0; Mismatches 1; Indels 2; Gaps 2;
 QY 62 GAGTACAGTGTATCCAGACAGACATCCCGGAGAGATCCGCTGCGCATCC 121
 DB 8533 GAGTACAGTGTATCCAGACAGACATCCCGGAGAGATCCGCTGCGCATCC 8592
 QY 122 TACCACACCGGAGCTGCTCTTCACTGTTCAACTGCTGAGCTGTGATCTGT 181
 DB 8593 TACCACACCGGAGCTGCTCTTCACTGTTCAACTGCTGAGCTGTGATCTGT 8652
 QY 182 GAGAGCCATGCGCCAGTGTGCGGCTTTGTGTACCAACCACTGAGACAGTGG 241
 DB 8653 GAGAGCCATGCGCCAGTGTGCGGCTTTGTGTACCAACCACTGAGACAGTGG 8712
 QY 242 CCAATGGGAGAGCCCTTCAAGGAGATGCGAGAGCTCTGAGGTTATATATCT 301
 DB 8713 CCAATGGGAGAGCCCTTCAAGGAGATGCGAGAGCTCTGAGGTTATATATCT 8772
 QY 302 GATCCCCCATGGAAGTCAAGAGGAGTGTGAGTGAAGAGATATACGTCTT 361
 DB 8773 GATCCCCCATGGAAGTCAAGAGGAGTGTGAGTGAAGAGATATACGTCTT 8832
 QY 362 CAAGGAGTCAATTAAGGAGAAATGCTTCTCCCAAGAAAGAAATCATCAGCCCTGT 421
 DB 8833 CAAGGAGTCAATTAAGGAGAAATGCTTCTCCCAAGAAAGAAATCATCAGCCCTGT 8892
 QY 422 TACCTCTACCTCTGCCCCCAGGTGCGAGCTGCTCTTTTCAAGATGATGAGCA 481
 DB 8893 TACCTCTACCTCTGCCCCCAGGTGCGAGCTGCTCTTTTCAAGATGATGAGCA 8952
 QY 482 AGTGTCTCTGATCCCAACAGACCAATATGTAATGCTCTGCTGAGCACTTATGAG 541
 DB 8953 AGTGTCTCTGATCCCAACAGACCAATATGTAATGCTCTGCTGAGCACTTATGAG 9011
 QY 542 GCTGGCTGACCACTGATCTATCTCTGACAGCTGGCTTGTGAGGAGTGAATG 601
 DB 9012 GCTGGCTGACCACTGATCTATCTCTGACAGCTGGCTTGTGAGGAGTGAATG 9071
 QY 602 CACTGGAGACATGATCACTGAGGAGACCCCTGAGAGAAAGTCAATCCAGACAG 661
 DB 9072 CACTGGAGACATGATCACTGAGGAGACCCCTGAGAGAAAGTCAATCCAGACAG 9131
 QY 662 ACAGATGTGACCAAGCAAAAGTGCATTAATGCAATGTTAAATGTGATTTACGAG 721
 DB 9132 ACAGATGTGACCAAGCAAAAGTGCATTAATGCAATGTTAAATGTGATTTACGAG 9191


```

OY 722 CTAGCTATGGAGCTGCTGCTCTCTAGTCCAGGAATCATGGGGTATGATCTGCTCTCCA 781
DB 9192 CTAGCTATGGAGCTGCTGCTCTCTAGTCCAGGAATCATGGGGTATGATCTGCTCTCCA 9251
OY 782 CCTGTGGAGCTGTAAAGCAAGCTCAGGCTTATCTCCCACTGGGGGCTGTGCCCCCTCG 841
DB 9252 CCTGTGGAGCTGTAAAGCAAGCTCAGGCTTATCTCCCACTGGGGGCTGTGCCCCCTCG 9310
OY 842 GGAAGCTTCCGTGGGAGAGCCCATCATCTGTGTTCAATAGTGTAGAGATCTAGCTAAAGCC 901
DB 9311 GGAAGCTTCCGTGGGAGAGCCCATCATCTGTGTTCAATAGTGTAGAGATCTAGCTAAAGCC 9370
OY 902 CCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 961
DB 9371 CCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 9430
OY 962 CGTGAAGTGTCTCTCAAGCTTAAAGCTGTGAAGCAAGAACTTGGGGGGGATGCTCCAGAT 1021
DB 9431 CGTGAAGTGTCTCTCAAGCTTAAAGCTGTGAAGCAAGAACTTGGGGGGGATGCTCCAGAT 9490
OY 1022 GTGGGATATCTGATCTGATCTGATCTGATCTGATCTGATCTGATCTGATCTGATCTGAT 1081
DB 9491 GTGGGATATCTGATCTGATCTGATCTGATCTGATCTGATCTGATCTGATCTGATCTGAT 9550
OY 1082 CCAAGCCAGAGGGGTCAAGGGGAGAGGTGCAACCTCAGCATGAGCCAAAGACTGGGGTCAAG 1141
DB 9551 CCAAGCCAGAGGGGTCAAGGGGAGAGGTGCAACCTCAGCATGAGCCAAAGACTGGGGTCAAG 9610
OY 1142 GAGCAGGTGTGTGTGAGCAGAGACTGGGGGCGGGGCGGGGCGGGGCGGGGCGGGGCGGGG 1201
DB 9611 GAGCAGGTGTGTGTGAGCAGAGACTGGGGGCGGGGCGGGGCGGGGCGGGGCGGGGCGGGG 9670
OY 1202 ATTTGCTTCAATGAAGCCCTCAAGAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCA 1261
DB 9671 ATTTGCTTCAATGAAGCCCTCAAGAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCA 9730
OY 1262 TGAATATCCAGAACTTTTGTACTTCTGTGTGTTAAATGTTTAAATTTTAAATTTTAA 1321
DB 9731 TGAATATCCAGAACTTTTGTACTTCTGTGTGTTAAATGTTTAAATTTTAAATTTTAA 9790
OY 1322 AAATAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 1370
DB 9791 AAATAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 9839

```

RESULT 3

AAH98172 standard; cDNA; 1846 BP.

AAH98172:

12-OCT-2001 (first entry)

Human EST-derived coding sequence SEQ ID NO: 29.

Human; sheep; pig; cow; fruit fly; Yeast; hamster; macaque; horse;
 tomato; monkey; dog; sea urchin; expressed sequence tag; EST;
 diagnostics; forensic test; gene mapping; genetic disorder;
 biodiversity; gene therapy; nutrition; ss.

Homo sapiens.

MO200154477-A2.

02-AUG-2001.

25-JAN-2001; 2001MO-US02687.

25-JAN-2000; 2000US-0491404.

17-JUL-2000; 2000US-0617746.

03-AUG-2000; 2000US-0631451.

-SEP-2000; 2000US-0663870.

```

XX (HYSE-) HYSEQ INC.
PA Tang YT, Liu C, Zhou P, Qian XB, Wang Z, Chen R, Aundi V;
XX Cao Y, Dermanac RA, Zhang J, Werhman T;
XX WPI; 2001-476164/51.
XX P-PSDB; AAM23513.
XX Isolated polypeptide for treatment of diseases, diagnostics, raising
PT antibodies and research use -
PS Claim 1; Page 202-203; 1275pp; English.
XX The present invention provides the protein and coding sequences of novel
CC proteins from a variety of organisms, including human, dog, cat, horse,
CC cow, pig, hamster, monkey, macaque, yeast, bacteria, fruit fly, sea
CC urchin and tomato. These were derived from expressed sequence tags (ESTs)
CC from the organism of interest. They can be used in diagnostics,
CC forensics, gene mapping, identification of mutations, to assess
CC biodiversity and for nutritional purposes. The present sequence is a cDNA
CC of the invention.
SQ Sequence 1846 BP; 397 A; 526 C; 538 G; 385 T; 0 other;

```

Query Match 63.6%; Score 888; DB 22; Length 1846;

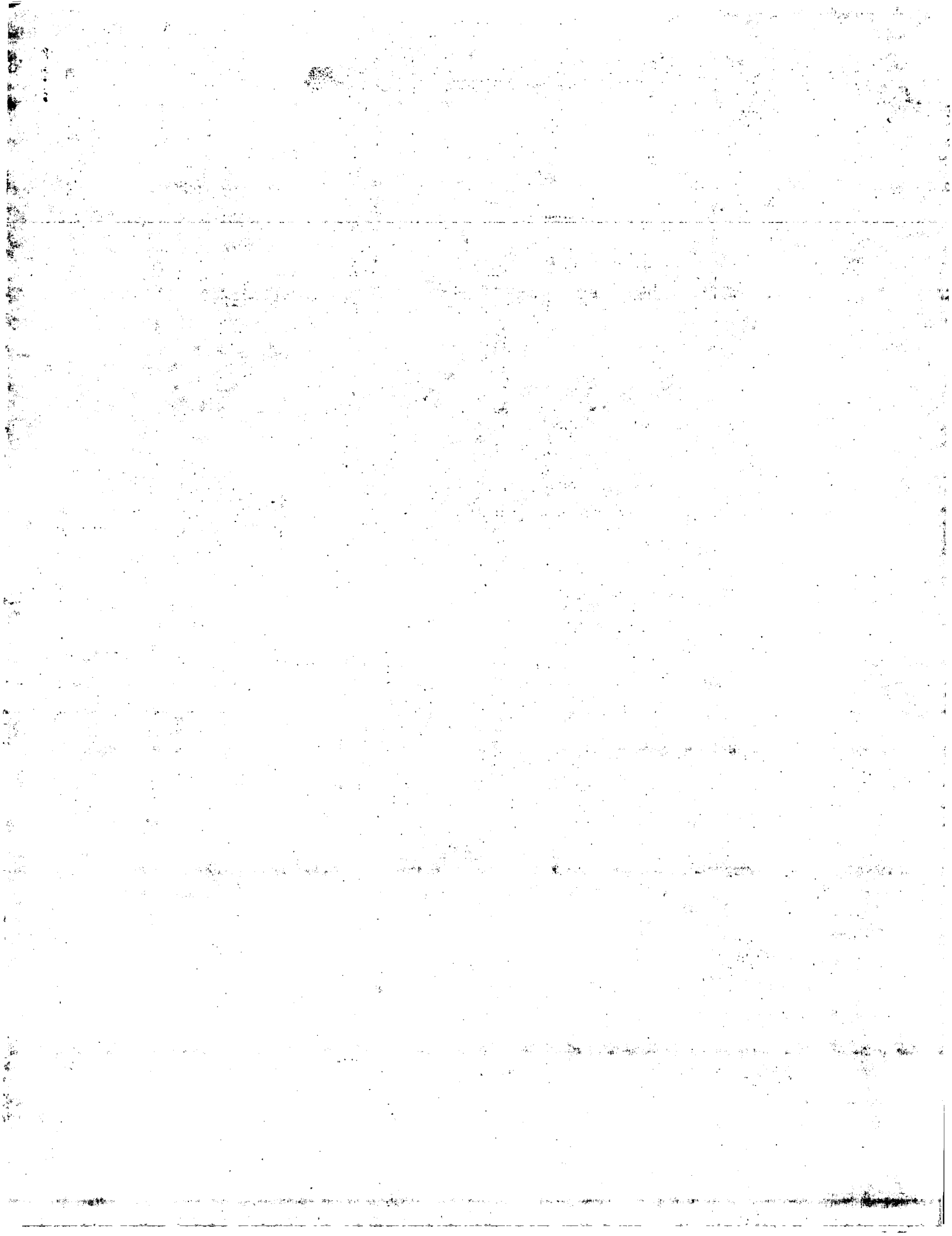
Best Local Similarity 99.9%; Pred. No. 0; Index 1; Gaps 0;

Matches 938; Conservative 0; Mismatches 1;

```

OY 442 CAGGTGGGAGCTGCTGCTTTTCAAGCTGAGTGAAGCAAGTGTGCTGATCCCAACA 501
DB 908 CAGGTGGGAGCTGCTGCTTTTCAAGCTGAGTGAAGCAAGTGTGCTGATCCCAACA 967
OY 502 AGACCAATATGTAAGGCTCTGCTGATCTGATCTGATCTGATCTGATCTGATCTGATCT 561
DB 968 AGACCAATATGTAAGGCTCTGCTGATCTGATCTGATCTGATCTGATCTGATCTGATCT 1027
OY 562 ATCTCAGAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 621
DB 1028 ATCTCAGAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1087
OY 622 ACTGGGAAAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 681
DB 1088 ACTGGGAAAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1147
OY 682 CGTGCAATATGCAAGTGTAAATGTAGTTTACAGCTGATGATGATGATGATGATGATGAT 741
DB 1148 CGTGCAATATGCAAGTGTAAATGTAGTTTACAGCTGATGATGATGATGATGATGATGAT 1207
OY 742 TCCTAGTCCAGGAATCATGGGGTATGACTGCTCTCAAGCTGCTGCTGCTGCTGCTGCTGCT 801
DB 1208 TCCTAGTCCAGGAATCATGGGGTATGACTGCTCTCAAGCTGCTGCTGCTGCTGCTGCTGCT 1267
OY 802 CTCAGGCTATGCTTCCCACTGGGGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 861
DB 1268 CTCAGGCTATGCTTCCCACTGGGGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1327
OY 862 CCATCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 921
DB 1328 CCATCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1387
OY 922 TGCACAGAGAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 981
DB 1388 TGCACAGAGAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1447
OY 982 TAGGCTGGAAGAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1041
DB 1448 TAGGCTGGAAGAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1507
OY 1042 GAGAGCTATCTGATCTGATCTGATCTGATCTGATCTGATCTGATCTGATCTGATCTGATCT 1101
DB 1508 GAGAGCTATCTGATCTGATCTGATCTGATCTGATCTGATCTGATCTGATCTGATCTGATCT 1567

```



SUMMARIES

score greater than or equal to the score of the result being printed and is derived by analysis of the total score distribution.

Result No.	Query Score	Length	DB	ID	Description
---------------	----------------	--------	----	----	-------------

Description:

AX056460 Sequence

AX480724 Séquence

AC122406 Mus muscu

AX056461 Sequence

AC117061 Ratūs no

AC121641 Rattus no

AC117177 Dictyoste

Continuation (2 of 2)

AC117075 Dictyoste

AX284646 Sequence

AC115603 Dictyoste

AC116305 Dictyoste

AX052773 Sequence

AP003869 Oryza sat

AP004654 Oryza sat

AC107116 Rattus no

AC116976 Dictyoste

AC115575 Dictyoste

AC116960 Dictyoste

AC115593 Dictyoste

AX284728 Sequence

—

10

PAT 12-AUG-2000

•

100

idae; Homo.

Sun, Y. and Liu, C.

JOURNAL Patent: WO 0242460-A 15 30-MAY-2002;
Diadex, Inc. (US)
Location/Qualifiers
1.1397
/organism="Homo sapiens"
/db_xref="taxon:9606"
BASE COUNT 332 a 363 c 398 g 304 t
ORIGIN

Query Match 100.0%; Score 1397; DB 6; Length 1397;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1397; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

1 GGTGTGACCTGTACCGGAGGCGAGATCTGAGAACTCCAGGGAAGAGAGATAC 60
1 GGTGTGACCTGTACCGGAGGCGAGATCTGAGAACTCCAGGGAAGAGAGATAC 60
61 CGAGTACCACTGTATCCAGAGACCAATCCCGAGAGAACTACCGCTCTGGCCATC 120
61 CGAGTACCACTGTATCCAGAGACCAATCCCGAGAGAACTACCGCTCTGGCCATC 120
121 CTACACACAGGAG 180
121 CTACACACAGGAG 180
181 TGAGAGCCATGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 240
181 TGAGAGCCATGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 240
241 GCGAGTGGAG 300
241 GCGAGTGGAG 300
301 TGATCCCGCATCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 360
301 TGATCCCGCATCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 360
361 TCAAGGAG 420
361 TCAAGGAG 420
421 TTACTCTACCTCTGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 480
421 TTACTCTACCTCTGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 480
481 AAGTGTCCCTGATCCCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 540
481 AAGTGTCCCTGATCCCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 540
541 GGGTGGGAG 600
541 GGGTGGGAG 600
601 GCACTGACAG 660
601 GCACTGACAG 660
661 GACAGATGTACAG 720
661 GACAGATGTACAG 720
721 CTTAGCTATGAG 780
721 CTTAGCTATGAG 780
781 ACCCTGTGGGCTGTAG 840
781 ACCCTGTGGGCTGTAG 840
841 GGGAGGCTTCCGAG 900
841 GGGAGGCTTCCGAG 900

901 CCTGTGCTGTGCTGTGCTGTGCTGTGCTGTGCTGTGCTGTGCTGTGCTGTGCT 960
901 CCTGTGCTGTGCTGTGCTGTGCTGTGCTGTGCTGTGCTGTGCTGTGCTGTGCT 960
961 TCGTGGAG 1020
961 TCGTGGAG 1020
1021 TGTGGAGATTTCTGATCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1080
1021 TGTGGAGATTTCTGATCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1080
1081 GCGAGCCAG 1140
1081 GCGAGCCAG 1140
1141 GAG 1200
1141 GAG 1200
1201 CATTTGCTTTCAATGAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1260
1201 CATTTGCTTTCAATGAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1260
1261 TTGAATATCCAGATTTTCTGATCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1320
1261 TTGAATATCCAGATTTTCTGATCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1320
1321 AAAATTAATTTAGTAAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1380
1321 AAAATTAATTTAGTAAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1380
1381 AAAATTAATTTAGTAAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1397
1381 AAAATTAATTTAGTAAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1397

RESULT 2
AC013480
LOCUS AC013480 128466 bp DNA linear PRI 07-NOV-2001
DEFINITION Homo sapiens BAC clone RP11-564016 from 2, complete sequence.
AC013480
VERSION AC013480.12 GI:16756369
KEYWORDS HTG.
SOURCE Homo sapiens.
ORGANISM Homo sapiens.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 128466).
Sulston, J.E. and Waterston, R.
Toward a complete human genome sequence
Genome Res. 8 (11), 1097-1108 (1998)
MEDLINE 99063792
PUBMED 9847074
REFERENCE 2 (bases 1 to 128466)
AUTHORS Nguyen, C. and Abbott, A.
TITLE The sequence of Homo sapiens BAC clone RP11-564016
JOURNAL Unpublished
REFERENCE 3 (bases 1 to 128466)
AUTHORS Waterston, R.H.
TITLE Direct Submission
JOURNAL Submitted (12-NOV-1999) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
REFERENCE 4 (bases 1 to 128466)
AUTHORS Waterston, R.
TITLE Direct Submission
JOURNAL Submitted (07-NOV-2001) Department of Genetics, Washington
University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
On Nov 7, 2001 this sequence version replaced gi:1339435.
COMMENT
Center: Washington University Genome Sequencing Center
Center code: WUGSC

```

/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="2"
/map="2"
/clone="RP11-564O16"
/clone_lib="RPCT-11"
291.452
repeat_region
/rpt_family="Alu"
1484.1617
/rpt_family="L2"
1669.1826
repeat_region
/rpt_family="Alu"
1864.2412
/rpt_family="MALR"
2413.2591
repeat_region
/rpt_family="MIR"
2689.2833
repeat_region
/rpt_family="L2"
3055.3082
repeat_region
/rpt_family="(T)n"
3085.3392
repeat_region
/rpt_family="Alu"
3530.3643
repeat_region
/rpt_family="CRI"
3644.3717
repeat_region

```

```

repeat_region      /rpt_family="Mariner"
                    3718..4183
repeat_region      /rpt_family="CRI"
                    6123..6293
repeat_region      /rpt_family="MIR"
                    6845..6881
repeat_region      /rpt_family="AT_rich"
                    6911..6963
repeat_region      /rpt_family="AT_rich"
                    7910..8008
repeat_region      /rpt_family="MIR"
                    8009..8299
repeat_region      /rpt_family="Alu"
                    8300..8326
repeat_region      /rpt_family="MIR"
                    8327..8599
repeat_region      /rpt_family="Alu"
                    8600..8668
repeat_region      /rpt_family="MIR"
                    9463..9553
repeat_region      /rpt_family="(TTCA)n"
                    9890..9923
repeat_region      /rpt_family="AT_rich"
                    10066..10473
repeat_region      /rpt_family="L2"
                    10474..10520
repeat_region      /rpt_family="(CA)n"
                    10521..10746
repeat_region      /rpt_family="L2"
                    12200..12385
repeat_region      /rpt_family="MIR"
                    12452..12489
repeat_region      /rpt_family="(CACG)n"
                    13410..13773
repeat_region      /rpt_family="MALR"
                    14076..14317
repeat_region      /rpt_family="MIR"
                    15470..15592
repeat_region      /rpt_family="MIR"
                    15599..15967
repeat_region      /rpt_family="MALR"
                    16227..16617
misc_feature       /note="similar to EST AA913068 (NID:93052460) ol34e06.s1"
repeat_region      16511..16693
repeat_region      /rpt_family="L1"
                    16717..16788
repeat_region      /rpt_family="L1"
                    16986..17009
repeat_region      /rpt_family="AT_rich"
                    18369..18509
repeat_region      /rpt_family="MER103"
                    18594..18644
repeat_region      /rpt_family="(TG)n"
                    19035..19193
repeat_region      /rpt_family="MER1_type"
                    19359..19668
repeat_region      /rpt_family="L2"
                    22454..22498
repeat_region      /rpt_family="GA-rich"
                    25334..25644
repeat_region      /rpt_family="Alu"
                    25647..25768
repeat_region      /rpt_family="MIR"
                    25770..25888
repeat_region      /rpt_family="L2"
                    25893..26193
misc_feature       /rpt_family="Alu"
                    25910..26358
misc_feature       /note="similar to EST AA578847 (NID:92357031) nk66a09.s1"
repeat_region      26366..26432
repeat_region      /rpt_family="MIR"
                    26502..26568
repeat_region      /rpt_family="MIR"

```

```

repeat_region 26729: .27274
/rpc_family="MALR"
repeat_region 27275: .27400
/rpc_family="MIR"
repeat_region 28344: .28444
/rpc_family="MIR"
repeat_region 28445: .28737
/rpc_family="Alu"
repeat_region 28738: .28877
/rpc_family="MIR"
repeat_region 29303: .29383
/rpc_family="Alu"
repeat_region 31059: .31158
/rpc_family="L1"
repeat_region 31180: .31624
/rpc_family="L1"
repeat_region 31626: .31773
/rpc_family="Alu"
repeat_region 31774: .31993
/rpc_family="Alu"

```

Query Match 93.7% Score 1309; DB 9; Length 128466;

Best Local Similarity 99.6%; Pred. No. 0;

Matches 1312; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

```

QY 59 ACCGATACCAAGTATCCAGACAGACACATCCCGAGAGACTACCGCTGCTGCCA 118
|
Db 113795 ACAGAGTACCAAGTATCCAGACAGACACATCCCGAGAGACTACCGCTG 113854
|
QY 119 TCTTACCAACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 178
|
Db 113855 TCTTACCAACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 113914
|
QY 179 TGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 238
|
Db 113915 TGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 113974
|
QY 239 GAGCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 298
|
Db 113975 GAGCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 114034
|
QY 299 AGTATATCCCATTCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 358
|
Db 114035 AGTATATCCCATTCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 114094
|
QY 359 CTTCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 418
|
Db 114095 CTTCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 114154
|
QY 419 TGTATCTCTACCTTCCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 478
|
Db 114155 TGTATCTCTACCTTCCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 114214
|
QY 479 CCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 538
|
Db 114215 CCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 114274
|
QY 539 AGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 598
|
Db 114275 AGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 114334
|
QY 599 TTGCACTGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 658
|
Db 114335 TTGCACTGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 114394
|
QY 659 CAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 718
|
Db 114395 CAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 114454
|
QY 719 AGCTATAGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 778
|
Db 114455 AGCTATAGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 114514
|

```

```

QY 779 CAACCTGTGGGCTGTAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 838
|
Db 114515 CAACCTGTGGGCTGTAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 114574
|
QY 839 CTGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 898
|
Db 114575 CTGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 114634
|
QY 899 GCCCTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 958
|
Db 114635 GCCCTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 114694
|
QY 959 CATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1018
|
Db 114695 CATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 114754
|
QY 1019 GATGTGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1078
|
Db 114755 GATGTGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 114814
|
QY 1079 AGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1138
|
Db 114815 AGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 114874
|
QY 1139 AGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1198
|
Db 114875 AGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 114934
|
QY 1199 CTCATTTGCTTCAATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1258
|
Db 114935 CTCATTTGCTTCAATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 114994
|
QY 1259 GTTGAATATCCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1318
|
Db 114995 GTTGAATATCCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 115054
|
QY 1319 ATTAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATA 1375
|
Db 115055 ATTAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATA 115111
|

```

RESULT 3
BC007901 1493 bp mRNA linear PRI 12-JUL-2001
LOCUS
DEFINITION
BC007901 Homo sapiens, clone IMAGE:4139786, mRNA, partial cds.
VERSION
BC007901.1 GI:14043927
KEYWORDS
SOURCE
ORGANISM
Homo sapiens.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 1493)
AUTHORS
Strausberg, R.
TITLE
Submitted (11-MAY-2001) National Institutes of Health, Mammalian
Gene Collection (MGC), Cancer Genomics Office, National Cancer
Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590.
USA

REMARK
COMMENT
NIH-MGC Project URL: <http://mgc.nci.nih.gov>
Contact: MGC help desk
Email: cgasbs-remail.nih.gov
Tissue Procurement: ATCC
CDNA Library Preparation: Rubin Laboratory
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (ILNL)
DNA Sequencing by: National Institutes of Health Intramural
Sequencing Center (NISC),
Gaithersburg, Maryland;
Web site: <http://www.nisc.nih.gov/>
Contact: nisc_mgc@nrci.nih.gov
Shevchenko, Y., Wetherby, K.D., Beckstrom-Sternberg, S.M.,
Benjamin, B., Blakesley, R.W., Bouffard, G.G., Brinkley, C., Brooks, S.,
Dietrich, N.L., Guan, X., Gupta, J., Ho, S.-L., Karlins, E., Legaspi, R.,

Lim, M., Maduro, Q. L., Masiello, C., Mastrian, S. D., McCloskey, J. C., McDowell, J., Pearson, R., Snyder, B., Stantrop, S., Thomas, P. J., Tongson, E. B., Touchman, J. W., Teurigne, C., Vogt, J. L., Walker, M. A., Zhang, L. H. and Green, E. D.

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/BLN at: <http://image.llnl.gov>
Series: IRAL Plate: 20 Row: 3 Column: 15
This clone was selected for full length sequencing because it passed the following selection criteria: Hexamer frequency ORF analysis.

FEATURES

Source Location/Qualifiers
1. 1493
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:4139786"
/tissue_type="Muscle, rhabdomyosarcoma"
/clone_id="NH_MGC_17"
/lab_host="DH10B-R"
/note="Vector: POT87"
c 1. 625
/codon_start=2
/product="Unknown (protein for IMAGE:4139786)"
/protein_id="AAH07901.1"
/db_xref="GI:14043928"
/translation="ANGELKVTLDLDARVEEPACSTDCILEFPANFTLPCSAQGM
CEGMEKRLNVAARFFPYLPHSAPELPLDSIVATGSLAGVETLAOLEKV
LHLYRGOYLONSTASSSTBYOCIPDSTIPDQBYRWPSPYHSHSCLSLSPNLEAVDV
CESHAOCRAFPVNTOTMTGRQLVFRKTMQVDPDPNKTYKASG"

CDS
BASE COUNT 360 a 401 c 422 g 310 t

Query Match 67.9%; Score 948; DB 9; Length 1493;
Best Local Similarity 84.3%; Pred. No. 7.1e-238;

Matches 112; Conservative 0; Mismatches 10; Indels 208; Gaps 1;

1 GGTGTCGACCTTACCGAGCGGAGATATCTGCAGAACTCCACGCGACGACGATAC 60
304 GGTGTCGACCTTACCGAGCGGAGATATCTGCAGAACTCCACGCGACGACGATAC 363
61 CGAGTACCACTGATCCCAACACGACCATCCCCGAGAAAGATACCGCTGCGCATC 120
364 CGAGTACCACTGATCCCAACACGACCATCCCCGAGAAAGATACCGCTGCGCATC 423
121 CTACCAACACGAGAGCTGCTCTTCAAGTTCACACCTGGCTGAGGCTGTGATGTC 180
424 CTACCAACACGAGAGCTGCTCTTCAAGTTCACACCTGGCTGAGGCTGTGATGTC 483
181 TGAGAGCACTGCCAGTGTGGGCTTTGTGTCAACACCAACACCACTGACAGGTGA 240
484 TGAGAGCACTGCCAGTGTGGGCTTTGTGTCAACACCAACCACTGACAGGTGA 536
241 GCCAGTGGAGAAACCTTTCAGAGGAGATGGACGACCTCTGTGAGGTTGATAGATG 300
537 ----- 536
301 TGATCCCCCATCGGAAGTCAAGAGGGGTCTGAGTGTATGAGAGAGATATACGTCT 360
537 ----- 536
361 TCAAGCAGTCAATTAGGAGAGATGTTGCTCCAGAAAGAAACATCCAGCCCTG 420
537 ----- 536
421 TTACTCTACCTCTGCCCCCAGGTGCGAGCTGCTTTTCAAGACTGATGAGCC 480
537 ----- 536
481 AAGTGTCCCTGATCCCAACACCAACATATGTAAGGCTCTGCTGACCTATCTGAG 540
576 AAGTGTCCCTGATCCCAACACCAACATATGTAAGGCTCTGCTGACCTATCTGAG 635
541 GGCTCGGCTGACCAAGTGTATCTCTGACAGCTGGGCTTGGCTGTGAGGAGATGACTT 600

|||||
Db 636 GGCTCGGCTGACCAAGTGTATCTCTGACAGCACTGGGCTTGGCTGTGAGGAGTACTT 695
601 GGAGTGGACCACTGATGTCACCTGGGAAACCTTCAGACAAAGTACATCCAGACA 660
696 GGAGTGGACCACTGATGTCACCTGGGAAACCTTCAGACAAAGTACATCCAGACA 755
661 GACAGATGTGACCAAGCAAAAGTGCATATATGCCAATGTTAAATGTGATGATCCAG 720
756 GACAGATGTGACCAAGCAAAAGTGCATATATGCCAATGTTAAATGTGATGATCCAG 815
721 CCTAGCTATGGAATCTGCTGCTCTTATGTCAGAGAAATCATGGGGATGACTGCTCTCA 780
816 CCTAGCTATGGAATCTGCTGCTCTTATGTCAGAGAAATCATGGGGATGACTGCTCTCA 875
781 ACCCTGTGGGCTGTATGACAAAGTCAAGGCTAGTCTCCCACTGGGGCTGTGCCCTCCCT 840
876 ACCCTGTGGGCTGTATGACAAAGTCAAGGCTAGTCTCCCACTGGGGCTGTGCCCTCCCT 935
841 GGAGCGGTTCCGTGGGACGCCCATCACTGTGTTCATATGATGATGATGATGATGATG 900
936 GGAGCGGTTCCGTGGGACGCCCATCACTGTGTTCATATGATGATGATGATGATGATG 995
901 CCTGTGCTGCTGTGACATGCAATGCAACAGAGCGGCTGGGCTGGGCAATGCA 960
996 CCTGTGCTGCTGTGACATGCAATGCAACAGAGCGGCTGGGCTGGGCAATGCA 1055
961 TGTGTAGTGTCTCTGACCTTATGATGATGATGATGATGATGATGATGATGATGATG 1020
1056 TGTGTAGTGTCTCTGACCTTATGATGATGATGATGATGATGATGATGATGATGATG 1115
1021 TGTGTAGTGTCTCTGACCTTATGATGATGATGATGATGATGATGATGATGATGATG 1080
1116 TGTGTAGTGTCTCTGACCTTATGATGATGATGATGATGATGATGATGATGATGATG 1175
1081 GCGAGCCAGGAGGTCAGAGGTCAGAGGTCAGACCTTACAGATGATGATGATGATGATG 1140
1176 GCGAGCCAGGAGGTCAGAGGTCAGAGGTCAGACCTTACAGATGATGATGATGATGATG 1235
1141 GGAGCGGTTCCGTGGGACGCCCATCACTGTGTTCATATGATGATGATGATGATGATG 1200
1236 GGAGCGGTTCCGTGGGACGCCCATCACTGTGTTCATATGATGATGATGATGATGATG 1295
1201 CATTTGCTTCAATGAAAGCTCAAAAGCAGCCAAACAGAGCTTCCCTTCTCGAGT 1260
1296 CATTTGCTTCAATGAAAGCTCAAAAGCAGCCAAACAGAGCTTCCCTTCTCGAGT 1355
1261 TGAATATCCAGATCTTTTGTACTTCTTGTGTTAAATGTTATTTTGTAAAAAT 1320
1356 TGAATATCCAGATCTTTTGTACTTCTTGTGTTAAATGTTATTTTGTAAAAAT 1415
1321 AAAATTAATTAATTAATTAATTAATGATGTTTCAACAGCAACTCTTCCCTTAAAAAAA 1380
1416 AAAATTAATTAATTAATTAATTAATGATGTTTCAACAGCAAAAAA 1475
1381 AAAAAAAAAA 1390
1476 AAAAAAAAAA 1485

RESULT 4
AX056460 2461 bp DNA linear PAT 13-JUN-2001
LOCUS AX056460
DEFINITION Sequence 104 from Patent W00073469.
ACCESSION AX056460
VERSION AX056460.1 GI:12229167
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
1 (bases 1 to 2461)
Plowman, G. D., Martinez, R., Whyte, D. and Sudersanam, S.

TITLE Protein Kinases
JOURNAL Patent: WO 0073469-A 104 07-DEC-2000;
Sugen, Inc. (US)
FEATURES Location/Qualifiers
Source 1. 2461
/organism="Homo sapiens"
/db_xref="taxon:9606"
BASE COUNT 426 a 771 c 811 g 453 t
ORIGIN

Query Match 67.3%; Score 940.4; DB 6; Length 2461;
Best Local Similarity 84.7%; Pred. No. 7.6e-236;
Matches 1159; Conservative 0; Mismatches 1; Indels 208; Gaps 1;

1 GGTCTGACCTGATGACCGGAGCGGAGTATCTGAGAACTCCAGCGGAGACAGTAC 60
1302 GGTCTGACCTGATGACCGGAGCGGAGTATCTGAGAACTCCAGCGGAGACAGTAC 1361

61 CGAGTACAGTGTATCCAGACAGACACATCCCGAGGAGACTACCGCTGCTGGCATC 120
1362 CGAGTACAGTGTATCCAGACAGACACATCCCGAGGAGACTACCGCTGCTGGCATC 1421

121 CTACACACAGCGGAGCTGCTCTTTAGTGTCAACCTGCTGAGCTGTGATGTCTG 180
1422 CTACACACACAGCGGAGCTGCTCTTTAGTGTCAACCTGCTGAGCTGTGATGTCTG 1481

181 TGAGAGCATGCGGAGTGTGGGCTTTGTGTGACCAACAGACCACTTGACAGAGTGA 240
1482 TGAGAGCATGCGGAGTGTGGGCTTTGTGTGACCAACAGACCACTTGAGAGTGA 1534

241 GCCAGTGGAGAAACCCCTTCCAGAGGAGATGGCAGAGCTCTCTGAGGTTGATAGTAG 300
1535 ----- 1534

301 TGATCCCCCATCGAAGTCAAGAGGGGCTGCTGAGGTATGAGAGAGTATACGTCT 360
1535 ----- 1534

361 TCAAGGAGTCAATTAGGAGAGATGCTTGTCTCCAGAAAGAAACATCCAGCCCTG 420
1535 ----- 1534

421 TTACCTCTACCTCTGCCCCCAGGTGCGAGCTGTCTTTTTCAGAGTGAATGAGCC 480
1535 ----- 1534

481 AAGTGTCCCTGATNCCCAAGACCAATATGAGAGGCTCTGCTGACCTATCTGAG 540
1574 AAGTGTCCCTGATNCCCAAGACCAATATGAGAGGCTCTGCTGACCTATCTGAG 1633

541 GGTCTGAGCTGACGCTGATCTCTCAGAGCTGAGGCTTGTGAGGAGAGTACTT 600
1634 GGTCTGAGCTGACGCTGATCTCTCAGAGCTGAGGCTTGTGAGGAGAGTACTT 1693

601 GCACTGCGACAGCTGATGTCACTCTGGGAAACCCCTGAGACAAAGCTTAACCTCCAGACA 660
1694 GCACTGCGACAGCTGATGTCACTCTGGGAAACCCCTGAGACAAAGCTTAACCTCCAGACA 1753

661 GACAGATGTGACAGAGCAAAAGCTGCAATATGCAAAATTTAAATGTGAGATTACAG 720
1754 GACAGATGTGACAGAGCAAAAGCTGCAATATGCAAAATTTAAATGTGAGATTACAG 1813

721 CCTAGCTATGGAAGTGTGCTCTTATGTCAGGAATCATGGGGTATGATGCTCTCA 780
1814 CCTAGCTATGGAAGTGTGCTCTTATGTCAGGAATCATGGGGTATGATGCTCTCA 1873

781 ACCCTGTGGCTTATGACAGCTCAGGCTATGCTCCCACTGGGGCTGTGCTTCCCT 840
1874 ACCCTGTGGCTTATGACAGCTCAGGCTATGCTCCCACTGGGGCTGTGCTTCCCT 1933

841 GGGAGGTTTCCGTGGGAGCCCATCATGTGTTCAATGTGAGAAATGTAGCTAAAGC 900
1934 GGGAGGTTTCCGTGGGAGCCCATCATGTGTTCAATGTGAGAAATGTAGCTAAAGC 1993

901 CCTCTGCTGCTGCTGCTGACATGCCACAGAGCGCGTGGGCTGCTGGGACAAATCCA 960
1994 CCTCTGCTGCTGCTGCTGCTGACATGCCACAGAGCGCGTGGGCTGCTGGGACAAATCCA 2053

961 TCGGAGGAGTGTCTCAGCTTATGAGTGTGAGACAGGAGACTTGGCGGGAGATGCTCCAGGA 1020
2054 TCGGAGGAGTGTCTCAGCTTATGAGTGTGAGACAGGAGACTTGGCGGGAGATGCTCCAGGA 2113

1021 TGTGGGTGATTTCTGATCCTGGGAGGCTATCTCTGACCTCCGACAGGGGACACTCCAG 1080
2114 TGTGGGTGATTTCTGATCCTGGGAGGCTATCTCTGACCTCCGACAGGGGACACTCCAG 2173

1081 GCCAGCCACAGGGGTGAGGGGACAGAGTGAACACTTACAGATGAGCCAAAGATGCGGCTAG 1140
2174 GCCAGCCACAGGGGTGAGGGGACAGAGTGAACACTTACAGATGAGCCAAAGATGCGGCTAG 2233

1141 GGAGCAGGTGTGTTGAGCCAGGACCTGGGGGCGGGGGTGGGGGCTTTCTGCTCT 1200
2234 GGAGCAGGTGTGTTGAGCCAGGACCTGGGGGCGGGGGTGGGGGCTTTCTGCTCT 2293

1201 CATTGCTTTCAATGAAAGCTCAAGACAGCCAAACAGAGCTTTCCCTTCTCTGAGT 1260
2294 CATTGCTTTCAATGAAAGCTCAAGACAGCCAAACAGAGCTTTCCCTTCTCTGAGT 2353

1261 TTGAATATCCAGAAATCTTTTGTACTCTTCTGTTGTTAAATTTGTTTAAATTAAT 1320
2354 TTGAATATCCAGAAATCTTTTGTACTCTTCTGTTGTTAAATTTGTTTAAATTAAT 2413

1321 AAAATTAATTTAGTTAATAATGATGTTTCAAGAGAACTCTTCCCT 1368
2414 AAAATTAATTTAGTTAATAATGATGTTTCAAGAGAACTCTTCCCT 2461

RESULT 5
AC112468
LOCUS
DEFINITION Rattus norvegicus clone CH230-12986, *** SEQUENCING IN PROGRESS
AC112468 94366 bp DNA linear HTG 17-JUL-2002
AC112468:2 GI:21745511
AC112468
KEYWORDS HTG; HTGS PHASE1.
SOURCE Norway rat.
ORGANISM Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
REFERENCE 1 (bases 1 to 94366)
Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-ouman,F.R., Allen,C., Albrooks,S.L., Amaralunge,H.C., Are,J.R., Ayele,M., Banks,T., Barbara,J., Benton,J., Bimaye,K., Blankenburg,K., Bonnin,D., Bouck,U., Bowie,S., Brileva,M., Brown,E., Brown,M., Bryant,N.P., Burch,P., Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C., Caron,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Z., Chowdhury,I., Christopoulos,C., Cleveland,C.D., Cox,C., Coyle,M.D., Dathorne,S.R., David,R., Davila,M.L., Davis,C., Day-Carroll,L., Dederich,D.A., Delaney,K.R., Delgado,O., Dem,A.L., Ding,Y., Dinh,H.H., Douthwaite,K.J., Draper,H., Dugan-Rocha,S., Durbin,K.J., Earhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Escoto,M., Falls,T., Ferraguto,D., Flagg,N., Ford,J., Foster,P., Frantz,P., Gabris,A., Gao,J., Garcia,A., Garner,T., Garza,N., Gill,R., Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S., Hamilton,K., Harris,C., Harris,K., Hart,M., Hayla,F., Hawes,A., Hernandez,J., Hernandez,O., Hodgson,A., Hughes,M., Holloway,C., Hollins,B., Homel,F., Howard,S., Huber,J., Huliy,S., Hume,J., Jackson,L.E., Jacobson,B., Jia,Y., Johnson,R., Jolivet,S., Joudah,S., Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J., Kovar,C., Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C., Lewis,L., Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,M., Louised,H., Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R., Ma,J., Maneshwari,M., Napua,P., Martin,R., Martindale,A., Martinez,E., Massey,E., Mawhiney,E., Mcleod,M.P., Meador,M., Mei,G., Metzker,M.,

TITLE
JOURNAL
REFERENCE
AUTHORS
JOURNAL
REFERENCE
AUTHORS
JOURNAL
COMMENT

Miner, G., Miner, Z., Mitchell, T., Mohabbat, K., Morgan, M., Morris, S., Moser, M., Neal, D., Newton, J., Newton, N., Nguyen, A., Nguyen, N., Nguyen, N., Nguyen, E., Nwokwen, S., Ogun, M., Okunola, G., Oragunye, N., Oviedo, R., Pace, A., Payton, B., Peery, J., Perez, L., Peters, L., Pichens, R., Primus, E., Pul, L., Quiles, M., Ren, Y., Rivas, M., Rojas, A., Rojibokan, I., Rolfe, M., Ruiz, S., Severy, G., Scherzer, S., Scott, G., Shen, H., Shoochari, N., Sisson, I., Sodergren, E., Sonalke, T., Sparks, A., Stanley, H., Stone, H., Sutton, A., Svatek, A., Tabor, P., Tamerisa, A., Thomas, N., Thomas, S., Tansy, J., Taylor, C., Taylor, T., Telford, B., Thomas, R., Wang, S., Wang, S., Ward-Moore, S., Warren, R., Washington, C., Watlington, S., Williams, G., Williamson, A., Wleczky, R., Wood, S., Worley, K., Wu, C., Wu, Y., Wu, Y., Zhou, J., Zorrilla, S., Nelson, D., Weinstein, G., and Gibbs, R.

Unpublished
2 (bases 1 to 94366)
Worley, K.C.
Direct Submission
Submitted (21-FEB-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
3 (bases 1 to 94366)
Worley, K.C.
Direct Submission
Submitted (17-JUL-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
On Jul 14, 2002 this sequence version replaced gi:18850446.

Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: <http://www.hgsc.bcm.tmc.edu/>
Contact: hgsc-help@bcm.tmc.edu
Project Information
Center project name: GS00
Center clone name: CH230-12986
Summary Statistics
Sequencing vector: Plasmid
Chemistry: Dye-terminator Big Dye 100% of reads
Assembly program: Phrap; version 0.990329
Consensus quality: 53937 bases at least Q40
Consensus quality: 56592 bases at least Q30
Consensus quality: 58737 bases at least Q20

NOTE: Estimated insert size may differ from sequence length
(see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
NOTE: This is a 'working draft' sequence. It currently
* consists of 44 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

1 1066: contig of 1066 bp in length
1067 1166: gap of unknown length
1167 2390: contig of 1224 bp in length
2391 2490: gap of unknown length
2491 3708: contig of 1218 bp in length
3709 3808: gap of unknown length
3809 5425: contig of 1617 bp in length
5426 5525: gap of unknown length
5526 6599: contig of 1074 bp in length
6599 6600: gap of unknown length
6600 8207: contig of 1508 bp in length
8208 8307: gap of unknown length
8308 9744: contig of 1437 bp in length
9745 9845: gap of unknown length
9845 11023: contig of 1179 bp in length
11024 11124: gap of unknown length
11124 12519: contig of 1396 bp in length

FEATURES
source
Location/Qualifiers
1. 94366
/organism="Rattus norvegicus"

12520 12619: gap of unknown length
12620 13673: contig of 1054 bp in length
13674 13773: gap of unknown length
13774 15620: contig of 1847 bp in length
15621 15720: gap of unknown length
15721 17466: contig of 1746 bp in length
17467 17566: gap of unknown length
17567 19136: contig of 1570 bp in length
19137 19236: gap of unknown length
19237 20973: contig of 1737 bp in length
20974 21073: gap of unknown length
21074 22091: contig of 1018 bp in length
22092 22191: gap of unknown length
22192 23778: contig of 1581 bp in length
23779 23878: gap of unknown length
23879 25293: contig of 1415 bp in length
25294 25393: gap of unknown length
25394 26983: contig of 1590 bp in length
26984 27083: gap of unknown length
27084 29408: contig of 2325 bp in length
29409 31243: gap of unknown length
31244 31343: contig of 1735 bp in length
31344 33604: gap of unknown length
33605 33704: gap of unknown length
33705 35038: contig of 1334 bp in length
35039 35138: gap of unknown length
35139 36399: contig of 1261 bp in length
36400 36499: gap of unknown length
36500 37623: contig of 1124 bp in length
37624 37723: gap of unknown length
37724 40099: contig of 2376 bp in length
40100 40199: gap of unknown length
40200 42674: contig of 2475 bp in length
42675 42774: gap of unknown length
42775 45617: contig of 2843 bp in length
45618 45717: gap of unknown length
45718 47425: contig of 1708 bp in length
47426 47525: gap of unknown length
47526 48809: contig of 1264 bp in length
48810 48909: gap of unknown length
48910 50692: contig of 1783 bp in length
50693 50792: gap of unknown length
50793 52171: contig of 1379 bp in length
52172 52271: gap of unknown length
52272 53845: contig of 1574 bp in length
53846 53945: gap of unknown length
53946 56501: contig of 2556 bp in length
56502 56601: gap of unknown length
56602 59622: contig of 3021 bp in length
59623 59722: gap of unknown length
59723 62615: contig of 2893 bp in length
62616 62715: gap of unknown length
62716 65623: contig of 2908 bp in length
65624 65723: gap of unknown length
65724 67513: contig of 1790 bp in length
67514 70311: gap of unknown length
70312 70411: contig of 2688 bp in length
70412 72769: contig of 2358 bp in length
72770 72869: gap of unknown length
72870 76193: contig of 3324 bp in length
76194 76293: gap of unknown length
76294 79434: contig of 3141 bp in length
79435 79534: gap of unknown length
79535 84177: contig of 4643 bp in length
84178 84277: gap of unknown length
84278 88578: contig of 4301 bp in length
88579 94366: gap of unknown length
94367 12519: contig of 5688 bp in length

BASE COUNT 22805 a 22369 c 21709 g 23135 t 4348 others
 ORIGIN

Query Match 35.6%; Score 497.2; DB 2; Length 94366;
 Best Local Similarity 70.3%; Pred. No. 3.8e-119;
 Matches 928; Conservative 0; Mismatches 303; Indels 89; Gaps 16;

60 CCGAGTACCAAGTATATCCAGACAGACATCCCGGAGAGATCACTACCGCTGCGCCAT 119
 Db 23973 CAGAGTACCAAGTATATCCAGACAGACATCCCGGAGAGATCACTACCGCTGCGCCAT 119
 Qy 120 CCTACCAAGTATATCCAGACAGACATCCCGGAGAGATCACTACCGCTGCGCCAT 119
 Db 24033 CCTACCAAGTATATCCAGACAGACATCCCGGAGAGATCACTACCGCTGCGCCAT 119
 Qy 180 GTGAGAGCATATCCAGACAGACATCCCGGAGAGATCACTACCGCTGCGCCAT 119
 Db 24093 GTGAGAGCATATCCAGACAGACATCCCGGAGAGATCACTACCGCTGCGCCAT 119
 Qy 240 AACCAGTGGAGAGATCCCGGAGAGATCACTACCGCTGCGCCAT 119
 Db 24153 AACCAGTGGAGAGATCCCGGAGAGATCACTACCGCTGCGCCAT 119
 Qy 300 GTGATCCCGGAGAGATCCCGGAGAGATCACTACCGCTGCGCCAT 119
 Db 24209 GTGATCCCGGAGAGATCCCGGAGAGATCACTACCGCTGCGCCAT 119
 Qy 359 GTGATCCCGGAGAGATCCCGGAGAGATCACTACCGCTGCGCCAT 119
 Db 24269 GTGATCCCGGAGAGATCCCGGAGAGATCACTACCGCTGCGCCAT 119
 Qy 415 GTGATCCCGGAGAGATCCCGGAGAGATCACTACCGCTGCGCCAT 119
 Db 24327 GTGATCCCGGAGAGATCCCGGAGAGATCACTACCGCTGCGCCAT 119
 Qy 475 GTGATCCCGGAGAGATCCCGGAGAGATCACTACCGCTGCGCCAT 119
 Db 24386 GTGATCCCGGAGAGATCCCGGAGAGATCACTACCGCTGCGCCAT 119
 Qy 535 GTGATCCCGGAGAGATCCCGGAGAGATCACTACCGCTGCGCCAT 119
 Db 24446 GTGATCCCGGAGAGATCCCGGAGAGATCACTACCGCTGCGCCAT 119
 Qy 595 GTGATCCCGGAGAGATCCCGGAGAGATCACTACCGCTGCGCCAT 119
 Db 24488 GTGATCCCGGAGAGATCCCGGAGAGATCACTACCGCTGCGCCAT 119
 Qy 655 GTGATCCCGGAGAGATCCCGGAGAGATCACTACCGCTGCGCCAT 119
 Db 24533 GTGATCCCGGAGAGATCCCGGAGAGATCACTACCGCTGCGCCAT 119
 Qy 715 GTGATCCCGGAGAGATCCCGGAGAGATCACTACCGCTGCGCCAT 119
 Db 24593 GTGATCCCGGAGAGATCCCGGAGAGATCACTACCGCTGCGCCAT 119
 Qy 771 GTGATCCCGGAGAGATCCCGGAGAGATCACTACCGCTGCGCCAT 119
 Db 24653 GTGATCCCGGAGAGATCCCGGAGAGATCACTACCGCTGCGCCAT 119
 Qy 831 GTGATCCCGGAGAGATCCCGGAGAGATCACTACCGCTGCGCCAT 119
 Db 24686 GTGATCCCGGAGAGATCCCGGAGAGATCACTACCGCTGCGCCAT 119
 Qy 891 GTGATCCCGGAGAGATCCCGGAGAGATCACTACCGCTGCGCCAT 119
 Db 24745 GTGATCCCGGAGAGATCCCGGAGAGATCACTACCGCTGCGCCAT 119
 Qy 951 GTGATCCCGGAGAGATCCCGGAGAGATCACTACCGCTGCGCCAT 119
 Db 24803 GTGATCCCGGAGAGATCCCGGAGAGATCACTACCGCTGCGCCAT 119

Qy 1011 TGCTCCAGAGATGAGGATTTCTGTACCTGAGAGATCACTACCGCTGCGCCAT 119
 Db 24860 TGCTCCAGAGATGAGGATTTCTGTACCTGAGAGATCACTACCGCTGCGCCAT 119
 Qy 1071 AACTCCAGAGATGAGGATTTCTGTACCTGAGAGATCACTACCGCTGCGCCAT 119
 Db 24917 AACTCCAGAGATGAGGATTTCTGTACCTGAGAGATCACTACCGCTGCGCCAT 119
 Qy 1131 CTGAGAGATGAGGATTTCTGTACCTGAGAGATCACTACCGCTGCGCCAT 119
 Db 24976 CTGAGAGATGAGGATTTCTGTACCTGAGAGATCACTACCGCTGCGCCAT 119
 Qy 1191 CTGAGAGATGAGGATTTCTGTACCTGAGAGATCACTACCGCTGCGCCAT 119
 Db 25036 CTGAGAGATGAGGATTTCTGTACCTGAGAGATCACTACCGCTGCGCCAT 119
 Qy 1251 TTCTCCAGAGATGAGGATTTCTGTACCTGAGAGATCACTACCGCTGCGCCAT 119
 Db 25096 TTCTCCAGAGATGAGGATTTCTGTACCTGAGAGATCACTACCGCTGCGCCAT 119
 Qy 1311 TGTAAAGATGAGGATTTCTGTACCTGAGAGATCACTACCGCTGCGCCAT 119
 Db 25155 TGTAAAGATGAGGATTTCTGTACCTGAGAGATCACTACCGCTGCGCCAT 119

RESULT 6
 AX480724 470 bp DNA linear PAT 12-AUG-2002
 LOCUS
 DEFINITION
 Sequence 14 from Patent WO0242460.
 AX480724
 VERSION
 AX480724.1 GI:22217466
 KEYWORDS
 SOURCE
 ORGANISM
 human.
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE
 AUTHORS
 Macina, R.A., Recipon, H., Pluta, J., Ghosh, M.G., Sun, Y., and Liu, C.
 TITLE
 Compositions and methods relating to colon specific genes and proteins
 JOURNAL
 Patent: WO 0242460-A 14 30-MAY-2002;
 Diadexus, Inc. (US)

FEATURES
 source
 location/Qualifiers
 1..470
 /organism="Homo sapiens"
 /db_xref="taxon:9606"

BASE COUNT 98 a 122 c 147 g 103 t
 ORIGIN

Query Match 33.2%; Score 464.2; DB 6; Length 470;
 Best Local Similarity 99.4%; Pred. No. 8e-111;
 Matches 466; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 567 CAGAGCTGGGCTTCTGAGAGAGATCACTACCGCTGCGCCAT 119
 Db 2 CAGAGCTGGGCTTCTGAGAGAGATCACTACCGCTGCGCCAT 119
 Qy 627 GGAACCCCTGAGAGATCACTACCGCTGCGCCAT 119
 Db 62 GGAACCCCTGAGAGATCACTACCGCTGCGCCAT 119
 Qy 687 AATAATGCCAAATGTTAAATGAGTTTACAGAGTATGAGAGATCACTACCGCTGCGCCAT 119
 Db 122 AATAATGCCAAATGTTAAATGAGTTTACAGAGTATGAGAGATCACTACCGCTGCGCCAT 119
 Qy 747 GTCCAGAGATCACTACCGCTGCGCCAT 119
 Db 182 GTCCAGAGATCACTACCGCTGCGCCAT 119
 Qy 807 GTTATGCTCCCACTGAGAGATCACTACCGCTGCGCCAT 119
 Db 242 GTTATGCTCCCACTGAGAGATCACTACCGCTGCGCCAT 119

Oy	867	ACTGTGTTCAATAAGTGTGAGAAATGTATGCTAAAGCCCTGCTGTGCTGTGCTGCACATAGCCA	926
Db	302	ACTGTGTTCAATAAGTGTGAGAAATGTATGCTAAAGCCCTGCTGTGCTGTGCTGCACATAGCCA	361
Oy	927	CAGCAGCGCGTGGGGCTGCTGCTGGGGACAATCATGCTGGAAGTTCCTCTCAGCTTAGT	986
Db	352	CAGCAGCGCGTGGGGCTGCTGCTGGGGACAATCATGCTGGAAGTTCCTCTCAGCTTAGT	421
Oy	987	CTGGAACAGAGACTTGGCGGGGGATCTCCAGGATGTGGGTGTAATCTGT	1035
Db	422	CTGGAACAGAGACTTGGCGGGGGATCTCCAGGATGTGGGTGTAATCTGT	470

LOCUS	AX408491/c			
DEFINITION	AX408491	427 bp	DNA	linear
ACCESSION	Sequence 1138 from Patent WO0229103.			
VERSION	AX408491			
KEYWORDS	AX408491.1 GI:21441196			
SOURCE	human.			
ORGANISM	Homo sapiens			
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.			
AUTHORS	Alvares, C., Horne, D., Peres-da-Silva, S. and Vockley, J.G.			
TITLE	Gene expression profiles in liver cancer			
JOURNAL	Patent: WO 0229103-A 1138 11-APR-2002;			
	GENE LOGIC INC (US)			

FEATURES	Location/Qualifiers
source	1..427
	/organism="Homo sapiens"
	/db_xref="taxon:9606"
	/note="EMBL/GenBank Accession No. AA451877"
BASE COUNT	107 a 126 c 92 g 102 t
ORIGIN	

Query Match	30.1%	Score 420.8	DB 6	Length 427
Best Local Similarity	99.5%	Pred. No. 2e-99		
Matches 422, Conservative	0	Mismatches 2	Indels 0	Gaps 0

QY	936	GTGGGGGCTGCCTGGGGACAATCATCTGTGAGTGTCTCTCAGCTTAGTCTGGACAG	995
Db	427	GTGGGGGCTGCCTGGGGACAATCATCTGTGAGTGTCTCTCAGCTTAGTCTGGACAG	368
QY	996	AGACTTGGCCGGGGGATTCCTCCAGGATGTGGGTGATTTCTGTACTCTGGGAGGCTAATCTCG	1055
Db	367	AGACTTGGCCGGGAGATGCTCCAGGATGTGGGTGATTTCTGTACTCTGGGAGGCTAATCTCTG	308
QY	1056	ACCTCCGACAGGGGACACTCCGACGACAGCCGAGGGGCTCAGGGGACAGGTGACACCT	1111
Db	307	ACCTCCGACAGGGGACACTCCGAGGCCAGCCGAGGGGCTCAGGGGACAGGTGACACCT	248
QY	1116	CAGCATGAGCCAAAGCTGGGGCTCAGGAGCAGGTGTGTTTATGACGACGACTCTGGGGCGG	1177
Db	247	CAGCATGAGCCAAAGCTGGGGCTCAGGAGCAGGTGTGTTTATGAGCAGGACTCTGGGGCGG	188
QY	1176	GGGTGGGGCCGGGGGCTTCTCTGCTCATTTTGCTTCAATGAAAGGCTCAAGCAGCCAAA	1235
Db	187	GGGTGGGACCGGGGCTTCTCTGCTCATTTTGCTTCAATGAAAGGCTCAAGCAGCCAAA	128
QY	1236	ACCAAGCTTTTCCCCCTTCTCTGAGTTGAATATCCAGAACTTTTGTACTTTTGTGTGT	1297
Db	127	ACCAAGCTTTTCCCCCTTCTCTGAGTTGAATATCCAGAACTTTTGTACTTTTGTGTGT	68
QY	1296	TAAATGTTTATTTTGTAAAAAATAAAATAAAATAGTAAATAAATGATGTTTTCACAG	1355
Db	67	TAAATGTTTATTTTGTAAAAAATAAAATAAAATAGTAAATAAATGATGTTTTCACAG	8
QY	1356	CAAA 1359	
Db	7	CAAA 4	

RESULT	8
ACI122406/c	
LOCUS	
DEFINITION	ACI122406 176771 bp DNA linear HTG 20-JUN-2002
VERSION	Mus musculus chromosome UNK clone RP24-126L14, WORKING DRAFT SEQUENCE, 12 unordered pieces.
KEYWORDS	ACI122406 ACI122406.2 GI:21490547
SOURCE	HTG; HTGS PHASE1; HTGS_DRAFT; HTGS_FULLTOP. house mouse.
ORGANISM	Mus musculus Eukaryota; Metazoa; Chordata; Craniota; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE	1 (bases 1 to 176771) McPherson,J.D. and Waterston,R.H.
AUTHORS	The sequence of Mus musculus clone Unpublished
TITLE	2 (bases 1 to 176771)
REFERENCE	McPherson,J.D. and Waterston,R.H.
AUTHORS	Direct Submission
TITLE	Submitted (23-MAY-2002) Genome Sequencing Center, 4444 Forest Park Parkway, St. Louis, MO 63108, USA
JOURNAL	3 (bases 1 to 176771) McPherson,J.D. and Waterston,R.H.
REFERENCE	Direct Submission
AUTHORS	Submitted (20-JUN-2002) Genome Sequencing Center, 4444 Forest Park Parkway, St. Louis, MO 63108, USA
TITLE	
JOURNAL	
COMMENT	On Jun 20, 2002 this sequence version replaced gi:21105866.

----- Genome Center -----
Center: Washington University Genome Sequencing Center
Center code: MUGSC
Web site: <http://genome.wustl.edu/gsc/index.shtml>
Contact: sdm@slonewatson.wustl.edu
Project Information
Center project name: M_BB012614

```
----- Summary Statistics -----
Sequencing vector: M13; 0%
Sequencing vector: plasmid; 100%
Chemistry: Dye-primer ET; 0% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.990319
Consensus quality: 171598 bases at least Q40
Consensus quality: 172211 bases at least Q30
Consensus quality: 172737 bases at least Q20
Insert size: 186000; agarose-fp
Insert size: 175614; sum-of-ctrls
Quality coverage: 11.87 in Q20 bases; agarose-fp
Quality coverage: 11.02 in Q20 bases; sum-of-ctrls
```

```
* NOTE: This is a 'working draft' sequence. It currently
* consists of 12 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
```

*	1	57:	contig of 57 bp in length
*	58	157:	gap of unknown length
*	158	3050:	contig of 2893 bp in length
*	3051	3150:	gap of unknown length
*	3151	5182:	contig of 2032 bp in length
*	5183	5282:	gap of unknown length
*	5283	14170:	contig of 8888 bp in length
*	14171	14270:	gap of unknown length
*	14271	22115:	contig of 7845 bp in length
*	22116	22215:	gap of unknown length
*	22216	32754:	contig of 10539 bp in length
*	32755	32854:	gap of unknown length
*	32855	46538:	contig of 13704 bp in length
*	46539	46558:	gap of unknown length

FEATURES
 source
 * 46559 65170: contig of 18512 bp in length
 * 65171 83892: contig of 18622 bp in length
 * 83893 83992: gap of unknown length
 * 83993 103925: contig of 19933 bp in length
 * 103926 104025: gap of unknown length
 * 104026 133284: contig of 29259 bp in length
 * 133285 133385: gap of unknown length
 * 133385 176771: contig of 43387 bp in length
 Location/Qualifiers
 1. 176771
 /organism="Mus musculus"
 /db_xref="taxon:10090"
 /chromosome="X"
 /clone="RP24-126L14"
 1. 57
 /note="assembly_name:Contig11"
 158. 3050
 /note="assembly_name:Contig15"
 3151. 5182
 /note="assembly_name:Contig16"
 5283. 14170
 /note="assembly_name:Contig17"
 14271. 22115
 /note="assembly_name:Contig18"
 22216. 32754
 /note="assembly_name:Contig19"
 32855. 46558
 /note="assembly_name:Contig20"
 46659. 65170
 /note="assembly_name:Contig21"
 65271. 83892
 /note="assembly_name:Contig22"
 83993. 103925
 /note="assembly_name:Contig23"
 104026. 133284
 /note="assembly_name:Contig24"
 133385. 176771
 /note="assembly_name:Contig25"

BASE COUNT 45173 a 42065 c 42632 g 45785 t 1116 others
 ORIGIN

Query Match 28.6%; Score 400; DB 2; Length 176771;
 Best Local Similarity 66.7%; Pred. No. 1.4e-93;
 Matches 925; Conservative 0; Mismatches 290; Indels 171; Gaps 18;

QY 62 GAGTACCAAGTATCCAGACAGACATCCAGAGAAAGTACCGCTGCGCATCC 121
 Db 57221 GAGTACCAAGTATCCAGACAGACATCCAGAGAAAGTACCGCTGCGCATCC 57162
 QY 122 TACCACCAAGGAGCTGCTTTCAGTGTCAACCTGAGCTGAGTGTCTGT 181
 Db 57161 TATCACCAAGGAGCTGCTTTCAGTGTCAACCTGAGCTGAGTGTCTGT 57102
 QY 182 GAGAGCATGCCAGTGTGGGCTTTGTGTCAACCAACCAACCACTGAGAGCTGAG 241
 Db 57101 GAGAGCATGCCAGTGTGGGCTTTGTGTCAACCAACCAACCACTGAGAGCTGAG 57042
 QY 242 CCAGTGGAGAGCCCTTCCAGAGAGATGAGAGCACTCTGAGAGTGTAGTACT 301
 Db 57041 CCAGTGGAGAGCCCTTCCAGAGAGATGAGAGCACTCTGAGAGTGTAGTACT 56983
 QY 302 GATCCCATCGAGAGTCAAGAGGGGTGTGAGTGTATGAGAGAGTATCGTGT 358
 Db 56982 CCCTCAGAAAGCCAGAGGGGTGTGAGTGTATGAGAGAGTATCGTGT 56923
 QY 359 -----CTTCAAGGAGTCAATTTAGGAGAAATGCTTGT 392
 Db 56922 TAT 56863
 QY 393 CCTCCAGAAAGAGAAATCATCCAGCTTCACTCACTCACTGAGAGAGAGAGAG 452
 Db 56862 CCTCCAGAAAGAGAAATCATCCAGCTTCACTCACTCACTGAGAGAGAGAGAG 56804

QY 453 CTGATCTTTTCAAGATGAGATGAGACCAAGTGTCTCTGATATCCCAACAGACATAT 512
 Db 56803 CTGATCTTTTCAAGATGAGATGAGACCAAGTGTCTCTGATATCCCAACAGACATAT 56744
 QY 513 GTGAAGGCTCTGAGTACCTATCTGAGAGCTGAGTACAGCTGATCTTCAAGAG 572
 Db 56743 GTGAAGGCTCTGAGTACCTATCTGAGAGCTGAGTACAGCTGATCTTCAAGAG 56702
 QY 573 CTGAGCTGCTGTGAGAGGAGTGTGACCTGAGAGACATGATGACCTGAGAGC 632
 Db 56701 CTGAGCTGCTGTGAGAGGAGTGTGACCTGAGAGACATGATGACCTGAGAGC 56656
 QY 633 CTGACAGCAAAAGTAAATCCAGACAGACAGATGTGACAGAGC-AAAGCTGCAATTA 691
 Db 56655 CTGACAGCAAAAGTAAATCCAGACAGACAGATGTGACAGAGC-AAAGCTGCAATTA 56596
 QY 692 TGCCAAATGTAAATGATGATTTACAGCTGATGAGTATGAGTGTGCTCTGATCA 751
 Db 56595 GCAAAATGTAAATGATGATTTACAGCTGATGAGTATGAGTGTGCTCTGATCA 56557
 QY 752 GGAATCATGGGGGTATGATGCTGCTCCAACTGAGGGGTGTAAGCA-GCTGAGGCTTA 810
 Db 56556 GGAATCATGGGGGTATGATGCTGCTCCAACTGAGGGGTGTAAGCA-GCTGAGGCTTA 56497
 QY 811 GTCTCCCACTGAGGGGTGTGCTGCTGCTGAGAGCGTTCGTTGAGAGCCCATCACTG 870
 Db 56496 GTCTCCCACTGAGGGGTGTGCTGCTGCTGAGAGCGTTCGTTGAGAGCCCATCACTG 56459
 QY 871 TGTCAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 930
 Db 56458 TGTCAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 56407
 QY 931 AGCGGTGAGGGGTGCTGAGGAGCAATCATGATGATGATGATGATGATGATGATGAT 990
 Db 56406 AGCGGTGAGGGGTGCTGAGGAGCAATCATGATGATGATGATGATGATGATGATGAT 56350
 QY 991 ACAGAGATGAGGGGGGTGCTGAGAGTGTGATGATGATGATGATGATGATGATGAT 1050
 Db 56349 ACAGAGATGAGGGGGGTGCTGAGAGTGTGATGATGATGATGATGATGATGATGAT 56292
 QY 1051 CTCTGACCTCCGAGAGGAGCACTCCAGGAGAGCCAGAGGGGTGAGGGGAGAGTGA 1110
 Db 56291 CTCTGACCTCCGAGAGGAGCACTCCAGGAGAGCCAGAGGGGTGAGGGGAGAGTGA 56232
 QY 1111 CACTGACATGAGCAAGTGTGAGGAGAGAGAGTGTGATGATGATGATGATGATGATGAT 1170
 Db 56231 CACTGACATGAGCAAGTGTGAGGAGAGAGAGTGTGATGATGATGATGATGATGATGAT 56173
 QY 1171 GCGGGGGGTGAGGGGGGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1230
 Db 56172 GCGGGGGGTGAGGGGGGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 56113
 QY 1231 CC-----AAACAGAGC 1242
 Db 56112 CCAGAGAGAGGCTCTCCCACTCCAGAGAGAGGCTCTCCCTCTCTGAGAGGAGAGC 56053
 QY 1243 TTTCCCTTCTGAGTGAATATCCAGATCTTTTGTATCTTGTGTTAAATG 1302
 Db 56052 TTTCCCTTCTGAGTGAATATCCAGATCTTTTGTATCTTGTGTTAAATG 55993
 QY 1303 TTTATTTTGTAAATAATTAATAATTAATAATTAATAATTAATAATTAATAATTAATAAT 1362
 Db 55992 TTTATTTTGTAAATAATTAATAATTAATAATTAATAATTAATAATTAATAATTAATAAT 55934
 QY 1363 TTTCCCT 1368
 Db 55933 TTTCCCT 55928

RESULT 9
 G21939/c
 LOCUS G21939 462 bp DNA linear STS 31-MAY-1996

DEFINITION human STS WI-15430, sequence tagged site.
 ACCESSION G21939
 VERSION G21939.1 GI:1342265
 STS: STS sequence; primer; sequence tagged site.
 KEYWORDS Homo sapiens STS derived from sequences in dbEST and the Unigene collection.
 SOURCE Homo sapiens
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 1 (bases 1 to 462)
 REFERENCE Hudson T.
 Whitehead Institute/MIT Center for Genome Research; Physically
 Mapped STS
 TITLE Whitehead Institute/MIT Center for Genome Research; Physically
 JOURNAL Mapped STS
 COMMENT Unpublished (1995)
 CONTACT: Thomas Hudson
 Whitehead Institute/MIT Center for Genome Research
 Whitehead Institute for Biomedical Research
 9 Cambridge Center, Cambridge MA 02142 USA
 Tel: 617 252 1900
 Fax: 617 252 1902
 Email: thudson@genome.wi.mit.edu
 Primer A: AGGGAAGATTGCTGTGAA
 Primer B: GCCTCAAGACAGCAAAAC
 STS size: 150
 PCR Profile:
 Presoak:
 Denaturation:
 Annealing: 56 degrees C
 Polymerization:
 PCR Cycles: 35
 Thermal Cycler:
 Protocol:
 Template: 10 ng
 Primer: each 5 pM
 dNTPs: each 4 mM
 Tag Polymerase: 0.025 units/ul
 Total Vol: 20 ul
 Buffer:
 MgCl2: 1.5 mM
 KCl: 50 mM
 Tris-HCl: 10 mM
 pH: 9.3
 Derived from dbEST (genbank accession R47986).
 Location/Qualifiers
 1. 462
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /map="192.9 CR from top of Chr2 linkage group"
 STS
 primer_bind 1. 150
 primer_bind 1. 20
 complement(132. 150)
 BASE COUNT 107 a 123 c 117 g 108 t 7 others
 ORIGIN
 Query Match 18.2%; Score 254.4; DB 11; Length 462;
 Best Local Similarity 94.5%; Pred. No. 1.1e-55;
 Matches 307; Conservative 0; Mismatches 12; Indels 6; Gaps 4;
 QY 1050 TCTCTACCTCCGAGAGGAGACATCCC--AGGCCAGCCAGGAGGTCAGGGGCAAGGT 1107
 DB 325 TCTGACCTTCCTCCGAGAGGAGACATCCCAGGCCAGCCAGGAGGTCAGGGGCAAGGT 266
 QY 1108 GCACA-CCTCAGCATGAG-CCAAGACTGGGTCAGGAGAGGTCGTGTTGAGCC--AG 1163
 DB 265 GCACACCTCAGCATGAGCCCAAGACTGGGTCAGGAGAGGTCGTGTTGAGCCAGG 206
 QY 1164 GACCTGGGGGGGGGGGGGGGGGGGGCTTTCTGCTCATTTGCTTCAATGAAGCCCTC 1223
 DB 205 ACCCTGGGGGGGGGGGGGGGGGGGGCTTTCTGCTCATTTGCTTCAATGAAGCCCTC 146

QY 1224 AAGACAGCAAAACAGAGCTTCCCTTCCTGAGTTGAATATCCAGATCTTTTGA 1283
 DB 145 AAGACAGCAAAACAGAGCTTCCCTTCCTGAGTTGAATATCCAGATCTTTTGA 86
 QY 1284 CTCTCTGTTGGTTAAATGTTTATTTTGTAAATAATTAATTAATTAATTAAT 1343
 DB 85 CTCTCTGTTGGTTAAATGTTTATTTTGTAAATAATTAATTAATTAATTAATTAAT 26
 QY 1344 GATGTTCAAGCAAACTCTCCCT 1368
 DB 25 GATGTTCAAGCAAACTCTCCCT 1
 RESULT 10
 LOCUS BC022157 1586 bp mRNA linear ROD 07-AUG-2002
 DEFINITION Mus musculus, clone IMAGE:5134400, mRNA, partial cds.
 ACCESSION BC022157
 VERSION BC022157.1 GI:18381057
 KEYWORDS house mouse.
 SOURCE Mus musculus
 ORGANISM Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 1 (bases 1 to 1586)
 REFERENCE Strausberg, R.
 Direct Submission
 Submitted (24-JAN-2002) National Institutes of Health, Mammalian
 Gene Collection (MGC), Cancer Genomics Office, National Cancer
 Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
 USA
 NIH-MGC Project URL: http://mgc.nci.nih.gov
 CONTACT: MGC help desk
 Email: cgabbs-remail.nih.gov
 Tissue Procurement: Jeffrey E. Green, M.D.
 CDNA Library Preparation: Life Technologies, Inc.
 DNA Sequencing by: Genome Sequence Centre,
 BC Cancer Agency, Vancouver, BC, Canada
 info@cgsc.bc.ca
 Steven Jones, Jennifer Asano, Ian Bosdet, Yaron Butterfield,
 Susanna Chan, Readman Chiu, Chris Fjell, Erin Garland, Ran Guin,
 Leticia Hsiao, Martin Krzywinski, Reta Kutsche, Oliver Lee, Soo
 Sen Lee, Victor Ling, Carrie Mathewson, Candice Mcleavy, Steven
 Ness, Pawan Pandon, Anna-Liisa Pirbhnu, Parvaneh Saeedi, Jacqueline
 Schein, Duane Smallue, Michael Smith, Lorraine Spence, Jeff Scott,
 Michael Thorne, Miranada Tsai, Natasja van den Bosch, Jill Vardy,
 George Yang, Scott Zuyderduyn, Marco Marra.
 Clone distribution: MGC clone distribution information can be found
 through the I.M.A.G.E. Consortium/LNL at: http://image.lnl.gov
 Series: IRAK Plate: 57 Row: m Column: 21
 This clone was selected for full length sequencing because it
 passed the following selection criteria: Hexamer frequency ORF
 analysis.
 Location/Qualifiers
 1. 1586
 /organism="Mus musculus"
 /db_xref="taxon:10090"
 /map="FVB/N"
 /clone="IMAGE:5134400"
 /issue_type="liver, normal, 5 month old male mouse."
 /clone_1ib="NCI CGAP_L19"
 /lab_host="DH10B"
 /note="Vector: PCMV-SPORT6"
 <1. 809
 /codon_start=3
 /product="Unknown (protein for IMAGE:5134400)"
 /protein_id="AAH22157.1"
 /db_xref="GI:18381058"
 /translation="IPDLTITTELGAPEVMIQLQTSNEDFRICLSIGRLHLHLAH
 SPLGVTLLDFRPROFVLVNGELKVTLDIDARVEETPTCSADCTLEFPAPRNFSPCS

ACGCGGNNKRNLYNARFFETVLLRHSAPSLRPLDSTVMTGELAMVDETLAC
 LETTAHLPRSSGYLONSTSSRAEYRIPDSATDPCWPSYHNGCLLSVFNIAEA
 IDVCSHAQCFRAVFNQITWTRKLVFRTGMNQVDPADKRTTYVAPG"

BASE COUNT 354 a 455 c 441 g 336 t
 ORIGIN

Query Match 18.0%; Score 251; DB 10; Length 1586;
 Best Local Similarity 67.1%; Pred. No. 1e-54;

Matches 649; Conservative 0; Mismatches 200; Indels 118; Gaps 16;

420 GTTACCTCTCACCCTTGCCTCCAGAGTGGAGCTGCTTTTTCAGAGCTGATGAGC 479
 699 GTACCAACACAGACCACTGACAGAGTGGAGCTGCTTTTTCAGAGCTGATGAGC 758
 480 CAAGTGTCTCTGATCCCAAGACCAATATGTAAGAGCTCTGCTGATCTGTA 539
 759 CAAGTGTCTCTGATCCCAAGACCAATATGTAAGAGCTCTGCTGATCTGTA 817
 540 GGGCTGGCTGACCAAGCTGATCTGAGAGCTGCTGCTGCTGAGAGAGTACT 599
 818 GGGCTGAGCTGAC-----CAGCTGGCTTGGCTGCTGAGAGAGTACT 860
 600 TGCAGTGCAGACATGATGATCCTGAGAACCCCTGAGACAAAGCTACCTCCAGAC 659
 861 -----TGATCCCACTGAGAACCCCTGAGACAAAGCTACCTCCAGAC 906
 660 AGACAGATGTGACCAAGAC-AAAGCTGATATATGCAAAATGTTAAATGATTTACC 718
 907 CAATGATGTACCAAGACCAAAAGCTGATATGCAAAATGTTAAATGATTTACC 966
 719 AGCTAGATGTGAGACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 778
 967 AGCTT-----CAGTCCAGACATGCTGTTGAGAACCCATGCTCTC 1005
 779 CAACCTGTGGGCTGTAGCA-AGCTGAGCTGATCTCTCCCACTGGGGCTGTGCTCTC 837
 1006 TGAAGCTGTAGCTGTAGCAAGGGCTCAGCTGCTCT-----TA 1044
 838 CTTGGAGAGCTTCCGTGGGAGAGCCCAATCACTGCTTCAATAGTGAAGATGATGTA 897
 1045 ACTGGAGAGCTTCCGTGGGAGAG-CCCAATGATGATGATGATGATGATGATGATG 1102
 898 AGCCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 957
 1103 AAC-----ACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1152
 958 CCAATGAGAGTGTCTCTGAGCTTGTGAGTGTGAGAGAGAGAGAGAGAGAGAGAGAG 1017
 1153 CAGTCTGAGAGTGTCTCTGAGCTTGTGAGTGTGAGAGAGAGAGAGAGAGAGAGAG 1211
 1018 GATGT 1077
 1212 -GATGT 1270
 1078 CAGGCCAGGCCAGGGGTCAAGGGGAGAGTGCACACCTGAGTGAAGCAAGCTGGGGT 1137
 1271 TAGGCCAGGCCAGGGGTCAAGGGGAGAGTGCATC-CATGAGCCAAAGCTGAGAT 1328
 1138 CAGGAGAGAGAGT 1197
 1329 GAGAGAGAGAGATGATGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1387
 1198 CCAATGATGT 1233
 1388 CCAATGATGT 1447
 1234 AAACAGAGGCTTCCCTCTCTGAGATGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1293
 1448 AGGCCAGAGGCTTCCCTCTCTGAGATGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1507
 1294 GTTAAATGT 1352
 1508 ATTAAATGT 1567

QY 1353 CAGCAAA 1359
 DB 1568 TGACAAA 1574

RESULT 11

AX056461

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITILE

JOURNAL

FEATURES

source

BASE COUNT

ORIGIN

Query Match

Best Local Similarity

Matches 644; Conservative

Score 235.8; DB 6; Length 1886;

Pred. No. 1e-50;

Mismatches 197; Indels 140; Gaps 15;

420 GTTACCTCTCACCCTTGCCTCCAGAGTGGAGCTGCTTTTTCAGAGCTGATGAGC 479

990 GTACCAACACAGACCACTGACAGAGTGGAGCTGCTTTTTCAGAGCTGATGAGC 1049

480 CAAGTGTCTCTGATCCCAAGACCAATATGTAAGAGCTCTGCTGATCTGTA 539

1050 CAAGTGTCTCTGATCCCAAGACCAATATGTAAGAGCTCTGCTGATCTGTA 1108

540 GGGCTGGCTGACCAAGCTGATCTGAGAGCTGCTGCTGCTGCTGCTGCTGCTGCT 599

1109 GGGCTGAGCTGAC-----CAGCTGGCTTGGCTGCTGAGAGAGTACT 1151

600 TGCAGTGCAGACATGATGATCCTGAGAACCCCTGAGACAAAGCTACCTCCAGAC 659

1152 -----TGATCCCACTGAGAACCCCTGAGACAAAGCTACCTCCAGAG 1197

660 AGACAGATGTGACCAAGAC-AAAGCTGATATATGCAAAATGTTAAATGATTTACC 718

1198 CAATGATGTACCAAGACCAAAAGCTGATATGCAAAATGTTAAATGATTTACC 1257

719 AGCTAGATGTGAGACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 778

1258 AGCTT-----CAGTCCAGACATGCTGTTGAGAACCCATGCTCTC 1296

779 CAACCTGTGGGCTGTAGCA-AGCTGAGCTGATCTCTCCCACTGGGGCTGTGCTCTC 837

1297 TGAAGCTGTAGCTGTAGAGAGGGCTCAGGCTGTCT-----TA 1335

838 CTTGGAGAGCTTCCGTGGGAGAGCCCAATCACTGCTTCAATAGTGAAGATGATGTA 897

1336 ACTGGAGAGCTTCCGTGGGAGAG-CCCAATGATGATGATGATGATGATGATGATG 1393

898 AGCCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 957

1394 AAC-----ACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1443

958 CCAATGAGAGTGTCTCTGAGCTTGTGAGTGTGAGAGAGAGAGAGAGAGAGAGAG 1017

1444 CAGTCTGAGAGTGTCTCTGAGCTTGTGAGTGTGAGAGAGAGAGAGAGAGAGAG 1501

*	13057	14625:	contig of 1569 bp in length
*	14626	14725:	gap of unknown length
*	14726	16071:	contig of 1346 bp in length
*	16072	16171:	gap of unknown length
*	16172	17358:	contig of 1187 bp in length
*	17359	17458:	gap of unknown length
*	17459	16497:	contig of 1039 bp in length
*	18448	16597:	gap of unknown length
*	18558	18637:	contig of 1040 bp in length
*	19638	19737:	gap of unknown length
*	19738	20912:	contig of 1175 bp in length
*	20913	21012:	gap of unknown length
*	21013	22076:	contig of 1064 bp in length
*	22077	22177:	gap of unknown length
*	22177	23303:	contig of 1127 bp in length
*	23304	23403:	gap of unknown length
*	23404	24483:	contig of 1080 bp in length
*	24484	24583:	gap of unknown length
*	24584	25662:	contig of 1079 bp in length
*	25663	25762:	gap of unknown length
*	25763	26770:	contig of 1208 bp in length
*	26971	27070:	gap of unknown length
*	27071	28260:	contig of 1190 bp in length
*	27071	28360:	gap of unknown length
*	28261	29603:	contig of 1243 bp in length
*	29604	29703:	gap of unknown length
*	29704	30722:	contig of 1019 bp in length
*	30723	30822:	gap of unknown length
*	30823	31869:	contig of 1047 bp in length
*	31870	31969:	gap of unknown length
*	31970	33087:	contig of 1118 bp in length
*	33088	33187:	gap of unknown length
*	33188	34332:	contig of 1045 bp in length
*	34333	34332:	gap of unknown length
*	34333	35611:	contig of 1279 bp in length
*	35612	35711:	gap of unknown length
*	35712	36623:	contig of 1112 bp in length
*	36624	36923:	gap of unknown length
*	36924	38144:	contig of 1221 bp in length
*	38145	38244:	gap of unknown length
*	38245	39645:	contig of 1401 bp in length
*	39646	39745:	gap of unknown length
*	39746	40774:	contig of 1029 bp in length
*	40775	40874:	gap of unknown length
*	40875	41958:	contig of 1084 bp in length
*	41959	42058:	gap of unknown length
*	42059	43446:	contig of 1388 bp in length
*	43447	43546:	gap of unknown length
*	43547	44870:	contig of 1324 bp in length
*	44871	44970:	gap of unknown length
*	44971	46204:	contig of 1234 bp in length
*	46205	46304:	gap of unknown length
*	46305	47377:	contig of 1073 bp in length
*	47378	47477:	gap of unknown length
*	47478	48744:	contig of 1267 bp in length
*	48745	48844:	gap of unknown length
*	48845	49891:	contig of 1047 bp in length
*	49892	49991:	gap of unknown length
*	49992	51319:	contig of 1328 bp in length
*	51320	51419:	gap of unknown length
*	51420	52502:	contig of 1083 bp in length
*	52503	52602:	gap of unknown length
*	52603	53755:	contig of 1153 bp in length
*	53756	53855:	gap of unknown length
*	53856	55330:	contig of 1465 bp in length
*	55321	55420:	gap of unknown length
*	55421	56482:	contig of 1062 bp in length
*	56483	56582:	gap of unknown length
*	56583	57759:	contig of 1177 bp in length
*	57760	57859:	gap of unknown length
*	57860	59002:	contig of 1143 bp in length
*	59003	59103:	gap of unknown length
*	59103	60205:	contig of 1103 bp in length

Peters, L., Pickens, R., Primus, E., Pu, L. L., Quiles, M., Ren, Y.,
Rives, M., Rojce, A., Rojce, A., Rojce, A., Rojce, A., Rojce, A.,
Scherer, S., Scott, G., Shen, H., Shoshitani, N., Sisson, I.,
Sodergren, E., Sonaike, T., Sparks, A., Stanley, H., Stone, H.,
Sutton, A., Svatek, A., Tabor, P., Tamerisa, A., Tamerisa, K., Tang, H.,
Tansy, J., Taylor, C., Taylor, T., Telford, B., Thomas, N., Thomas, S.,
Umani, K., Vasquez, L., Vera, V., Villalon, D., Vinson, R., Wang, Q.,
Wang, S., Ward-Moore, S., Warren, R., Washington, C., Wallington, S.,
Williams, G., Williamson, A., Wleczky, R., Wooden, S., Worley, K.,
Wu, C., Wu, Y., Wu, Y. F., Zhou, J., Zorrilla, S., Nelson, D.,
Weinstock, G., and Gibbs, R.

Unpublished
Direct Submission
2 (bases 1 to 200207)
Worley, K.C.

Submitted (06-APR-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
3 (bases 1 to 200207)

Submitted (18-JUL-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
On Jul 14, 2002 this sequence version replaced gi:20066112.

Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: <http://www.hgsc.bcm.tmc.edu/>
Contact: hgsc-help@bcm.tmc.edu
Project Information
Center project name: GUSK
Center clone name: CH230-403121
Summary Statistics
Sequencing vector: Plasmid;
Chemistry: Dye-terminator Big Dye 100% of reads
Assembly program: Phrap; version 0.990329
Consensus quality: 66669 bases at least Q40
Consensus quality: 70534 bases at least Q30
Consensus quality: 73881 bases at least Q20

NOTE: Estimated insert size may differ from sequence length
* (see http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 112 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

1 1307: contig of 1307 bp in length
1308 1407: gap of unknown length
1408 2750: contig of 1343 bp in length
2751 2850: gap of unknown length
2851 4107: contig of 1257 bp in length
4108 4207: gap of unknown length
4208 5462: contig of 1255 bp in length
5463 5562: gap of unknown length
5563 6941: contig of 1379 bp in length
6942 7041: gap of unknown length
7042 8325: contig of 1284 bp in length
8326 8425: gap of unknown length
8426 9428: contig of 1003 bp in length
9429 9528: gap of unknown length
9529 10838: contig of 1310 bp in length
10839 12015: contig of 1077 bp in length
12016 12115: gap of unknown length
12116 13380: contig of 1265 bp in length
13381 13480: gap of unknown length
13481 14483: contig of 1003 bp in length

14484 14583: gap of unknown length
14584 15814: contig of 1231 bp in length
15815 15914: gap of unknown length
15915 16914: contig of 1000 bp in length
16915 17014: gap of unknown length
17015 18252: contig of 1238 bp in length
18253 18352: gap of unknown length
18353 19639: contig of 1287 bp in length
19640 19739: gap of unknown length
19740 21007: contig of 1268 bp in length
21008 21107: gap of unknown length
21108 22132: contig of 1025 bp in length
22133 22232: gap of unknown length
22233 23307: contig of 1075 bp in length
23308 23407: gap of unknown length
23408 24549: contig of 1142 bp in length
24550 24649: gap of unknown length
24650 26033: contig of 1384 bp in length
26034 26133: gap of unknown length
26134 27327: contig of 1194 bp in length
27328 27427: gap of unknown length
27428 28447: contig of 1020 bp in length
28448 28547: gap of unknown length
28548 30342: contig of 1795 bp in length
30343 31486: contig of 1044 bp in length
31487 31586: gap of unknown length
31587 32858: contig of 1272 bp in length
32859 32958: gap of unknown length
32959 33977: contig of 1019 bp in length
33978 34077: gap of unknown length
34078 35589: contig of 1512 bp in length
35590 35689: gap of unknown length
35690 37047: contig of 1358 bp in length
37048 37147: gap of unknown length
37148 38713: contig of 1566 bp in length
38714 38813: gap of unknown length
38814 39996: contig of 1183 bp in length
39997 40096: gap of unknown length
40097 41126: contig of 1030 bp in length
41127 41226: gap of unknown length
41227 42462: contig of 1236 bp in length
42463 42562: gap of unknown length
42563 43937: contig of 1375 bp in length
43938 44037: gap of unknown length
44038 45068: contig of 1031 bp in length
45069 45168: gap of unknown length
45169 46348: contig of 1180 bp in length
46349 46448: gap of unknown length
46450 47650: contig of 1202 bp in length
47651 47750: gap of unknown length
47751 49382: contig of 1632 bp in length
49383 49482: gap of unknown length
49484 50884: contig of 1402 bp in length
50885 50984: gap of unknown length
50985 52017: contig of 1033 bp in length
52018 52117: gap of unknown length
52118 53401: contig of 1284 bp in length
53402 53501: gap of unknown length
53502 54545: contig of 1044 bp in length
54546 54645: gap of unknown length
54646 55789: contig of 1144 bp in length
55790 55889: gap of unknown length
55890 56979: contig of 1090 bp in length
56980 57079: gap of unknown length
57080 58407: contig of 1328 bp in length
58408 58507: gap of unknown length
58509 59633: contig of 1126 bp in length
59634 59733: gap of unknown length
59734 60812: contig of 1079 bp in length
60813 61915: gap of unknown length
61916 62015: contig of 1003 bp in length

BASE COUNT 50753 a 15240 c 15399 g 51959 t 150 others
ORIGIN
Query Match 3.8%; Score 53.2; DB 2; Length 133501;
Best Local Similarity 68.9%; Pred. No. 0.018;
Matches 73; Conservative 0; Mismatches 33; Indels 0; Gaps 0;
QY 1291 TTGGTAAATGTTTATTTTGTATATAAATAAATTAATGTAATAATGATGTTT 1350
DB 92088 TTTTGTGAATGTTTATGATACAAATAAATAAATAAATAAATAAATAAATTT 92029
QY 1351 CACGAACTCTTCCCTAAATAAATAAATAAATAAATAAAGCGCT 1396
DB 92028 AAAAGTGCAATGTTTGTATATAAATAAATAAATAAATAAAGTGAAT 91983
RESULT 15
AC121641/c
LOCUS
DEFINITION
Rattus norvegicus clone CH230-435023, *** SEQUENCING IN PROGRESS
AC121641
AC121641.2 GI:21903291
VERSION
HTG; HTGS PHASE1.
KEYWORDS
Rattus norvegicus
SOURCE
Rattus norvegicus
ORGANISM
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
1 (bases 1 to 158007)
REFERENCE
AUTHORS
Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C.,
Alebrooks,S.L., Amaralunga,H.C., Are,J.R., Ayele,M., Banks,T.,
Barbier,U., Benton,J., Bimaga,K., Blankenburg,K., Bonini,D.,
Barbier,U., Bowler,S., Brileva,M., Brown,E., Brown,M., Bryant,N.P.,
Bouchay,C., Burich,P., Burkett,C., Butrell,K.L., Byrd,N.C.,
Carroll,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D.,
Chen,G., Chen,R., Chen,Z., Chowdhry,I., Christopoulos,C.,
Cleveland,C.D., Cox,C., Coyle,M.D., Dathorne,S.R., David,R.,
Davila,M.L., Davis,C., Davy-Carroll,L., Dederich,D.A.,
Delaney,K.R., Delgado,O., Denu,A.L., Ding,Y., Dinh,H.H.,
Douthwaite,K.U., Draper,H., Dugan-Rocha,S., Dudin,K.U.,
Barnhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Escoto,M.,
Falls,T., Ferraguto,D., Flagg,N., Ford,J., Foster,P., Frantz,P.,
Gabisi,A., Gao,J., Garcia,A., Garner,T., Garza,N., Gill,R.,
Gorrell,J.H., Guevara,M., Gunaratne,P., Hale,S., Hamilton,K.,
Harris,C., Harris,K., Hart,M., Haylak,P., Hawes,A., Hernandez,J.,
Hernandez,O., Hodgson,A., Hogues,M., Holloway,C., Hollins,B.,
Homs,F., Howard,S., Huber,J., Huiyk,S., Hume,J., Jackson,L.E.,
Jacobson,B., Jia,Y., Johnson,R., Jolivet,S., Joudah,S.,
Kratovic,J., Kurishi,A., Khan,U., King,L., Korvah,J., Kovar,C.,
Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J., Kovar,C.,
Li,U., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W., Louised,H.,
Lozad,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R., Ma,J.,
Maheshwari,M., Mapua,P., Martin,R., Martindale,A., Martinez,E.,
Massey,E., Mawlin,E., McLeod,M.P., Meador,M., Mei,G., Metzger,M.,
Miner,G., Miner,Z., Mitchell,T., Mohabakat,K., Morgan,M., Morris,S.,
Moser,M., Neal,D., Newton,J., Newton,S., Nguyen,A., Nguyen,N.,
Nguyen,N., Nicholson,E., Nwokenko,S., Oguh,M., Okunom,G.,
Ogunyeye,N., Oviedo,R., Pace,A., Payton,B., Peery,J., Perez,L.,
Peters,L., Pichens,R., Primus,E., Pu,L.L., Quilis,M., Ren,Y.G.,
Rives,M., Rojas,A., Rojupokan,I., Rolfe,M., Ruiz,S., Severy,G.,
Scherer,S., Scott,G., Shen,H., Shoshitari,N., Sisson,I.,
Sodergren,E., Sonaike,T., Sparks,A., Stanley,H., Stone,H.,
Sutton,A., Svatek,A., Taylor,T., Telford,B., Thomas,N., Thomas,S.,
Tansley,J., Taylor,C., Taylor,T., Telford,B., Thomas,N., Thomas,S.,
Tang,S., Ward-Moore,S., Warren,R., Washington,C., Watlington,C.,
Williams,G., Williamson,A., Wleczky,R., Woodson,S., Worley,K.,
Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D.,
Weinstock,G., and Gibbs,R.
TITLE
JOURNAL
Unpublished
Direct Submission
1 (bases 1 to 158007)

RESULT 14
AC116956/c
LOCUS
DEFINITION
Dictyostelium discoideum chromosome 2 map 1512060-1645559 strain
AX4, *** SEQUENCING IN PROGRESS *** in ordered pieces.
AC116956
AC116956.1 GI:19919988
VERSION
HTG; HTGS PHASE2.
KEYWORDS
Dictyostelium discoideum.
SOURCE
Dictyostelium discoideum
ORGANISM
Eukaryota; Mycetozoa; Dictyostelida; Dictyostelium.
1 (bases 1 to 133501)
REFERENCE
AUTHORS
Gloeckner,G., Eichinger,L., Szafranski,K., Pachob,J., Dear,P.,
Lemman,R., Baumgart,C., Parra,G., April,J.F., Guigo,R., Kumpf,K.,
Tunggal,B., Cox,E., Quail,M.A., Platzer,M., Rosenthal,A., and
Noegel,A.A.
TITLE
JOURNAL
Unpublished
Direct Submission
The Dictyostelium Genome Sequencing Consortium
2 (bases 1 to 133501)
REFERENCE
AUTHORS
Baumgart,C.
TITLE
JOURNAL
Submitted (04-Apr-2002) Genome Analysis, Institute of Molecular
Biotechnology, Beutenbergstr. 11, Jena 07745, Germany
CDS predictions from Genedit may contain errors. Further information
is available from IMB Jena, Department of Genome Analysis
(http://genome.imb-jena.de/dictyostelium/)
and the University Cologne, Institute for Biochemistry I
(http://www.uni-koeln.de/dictyostelium/project.shml)
Funding
Agency : Deutsche Forschungsgemeinschaft (DFG) :
* NOTE: This is a working draft sequence.
* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.
location/Qualifiers
1. 133501
/organism="Dictyostelium discoideum"
/strain="AX4"
/db_xref="taxon:44689"
/chromosome="2"
/map="1512060-1645559"

AUTHORS
TITLE
JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

Worley, K.C.
Direct Submission
Submitted (21-MAY-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
3 (bases 1 to 158007)
Worley, K.C.
Direct Submission
Submitted (23-JUL-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
On Jul 18, 2002 this sequence version replaced gi:21039560.

----- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: <http://www.hgsc.bcm.tmc.edu/>
Contact: hgsc-help@bcm.tmc.edu
----- Project Information
Center project name: GWC2
Center clone name: CH230-435D23
----- Summary Statistics
Sequencing vector: Plasmid
Chemistry: Dye-terminator Big Dye 100% of reads
Assembly program: Phrap, version 0.990329
Consensus quality: 101786 bases at least Q40
Consensus quality: 105464 bases at least Q30
Consensus quality: 108013 bases at least Q20

----- NOTE: Estimated insert size may differ from sequence length
* (see http://www.hgsc.bcm.tmc.edu/docs/Genbankdraft_data.html).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 55 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

1 1143: contig of 1143 bp in length
1144 1243: gap of unknown length
1244 2605: contig of 1362 bp in length
2606 2705: gap of unknown length
2706 3983: contig of 1278 bp in length
3984 4084: gap of unknown length
4084 5429: contig of 1346 bp in length
5430 5529: gap of unknown length
5530 6913: contig of 1384 bp in length
6914 7013: gap of unknown length
7014 8256: contig of 1243 bp in length
8257 8357: gap of unknown length
8357 9498: contig of 1142 bp in length
9499 9598: gap of unknown length
9599 10637: contig of 1039 bp in length
10638 10737: gap of unknown length
10738 11892: contig of 1155 bp in length
11893 11992: gap of unknown length
11993 13170: contig of 1178 bp in length
13171 13270: gap of unknown length
13271 14303: contig of 1033 bp in length
14304 14403: gap of unknown length
14404 15701: contig of 1298 bp in length
15702 15801: gap of unknown length
15802 16901: contig of 1100 bp in length
16902 17001: gap of unknown length
17002 18052: contig of 1051 bp in length
18053 18152: gap of unknown length
18153 19776: contig of 1624 bp in length
19777 19876: gap of unknown length
19877 21143: contig of 1267 bp in length
21144 21243: gap of unknown length
21244 22441: contig of 1198 bp in length
22442 22541: gap of unknown length
22542 24034: contig of 1493 bp in length

24035 24134: gap of unknown length
24135 25621: contig of 1487 bp in length
25622 25721: gap of unknown length
25722 27224: contig of 1503 bp in length
27225 27324: gap of unknown length
27325 29327: contig of 1903 bp in length
29328 29327: gap of unknown length
30598 30598: contig of 1271 bp in length
30599 30698: gap of unknown length
30699 32237: contig of 1539 bp in length
32238 32337: gap of unknown length
32338 33956: contig of 1619 bp in length
33957 34056: gap of unknown length
34057 35436: contig of 1380 bp in length
35437 35536: gap of unknown length
35537 37417: contig of 1881 bp in length
37418 37517: gap of unknown length
37518 39147: contig of 1630 bp in length
39148 39247: gap of unknown length
39248 40937: contig of 1690 bp in length
40937 41037: gap of unknown length
41038 43441: contig of 2404 bp in length
43442 43541: gap of unknown length
43542 45283: contig of 1742 bp in length
45284 45383: gap of unknown length
45384 47906: contig of 2523 bp in length
47906 48006: gap of unknown length
48007 50783: contig of 2777 bp in length
50784 50883: gap of unknown length
50884 53210: contig of 2327 bp in length
53211 53310: gap of unknown length
53311 56392: contig of 3082 bp in length
56393 56493: gap of unknown length
56493 59673: contig of 3181 bp in length
59673 59773: gap of unknown length
59773 62824: contig of 3051 bp in length
62824 62924: gap of unknown length
62924 66675: contig of 3751 bp in length
66675 66775: gap of unknown length
66775 70074: contig of 3299 bp in length
70074 70174: gap of unknown length
70175 72804: contig of 2630 bp in length
72805 72904: gap of unknown length
72905 76454: contig of 3550 bp in length
76455 76554: gap of unknown length
76555 80449: contig of 3895 bp in length
80450 80549: gap of unknown length
80550 83438: contig of 2889 bp in length
83439 83539: gap of unknown length
83539 88068: contig of 4530 bp in length
88069 88169: gap of unknown length
88169 93861: contig of 5693 bp in length
93862 93961: gap of unknown length
93962 98834: contig of 4873 bp in length
98835 98934: gap of unknown length
98935 102910: contig of 3976 bp in length
102911 103010: gap of unknown length
103011 106942: contig of 3932 bp in length
106943 107042: gap of unknown length
107043 110510: contig of 3468 bp in length
110511 110610: gap of unknown length
110611 114467: contig of 3857 bp in length
114468 114567: gap of unknown length
114568 118093: contig of 3526 bp in length
118094 118193: gap of unknown length
118194 125078: contig of 6885 bp in length
125079 125178: gap of unknown length
125179 130442: contig of 5264 bp in length
130443 130542: gap of unknown length
130543 136904: contig of 6362 bp in length
136905 137004: gap of unknown length
137005 143575: contig of 6571 bp in length

Query Match 3.8; Score 52.8; DB 2; Length 15807;
Best Local Similarity 51.3; Pred. No. 0.024;
Matches 117; Conservative 0; Mismatches 111; Indels 0; Gaps 0;

QY	1169	GGGGGGGGGGTGGGGCCGGGGCTTTCTGCTCATTTGCTTCATGAAAGCTCAAGC	1228
Db	2281	GGGGGGGGGGGGGGGGGGTTTTCCTCCNCCCCCCCCCCCCGGGGGGGGGG	2222
QY	1229	AGCCAAACGAGGCTTTCCCTTCCTCGAGTTGAATATCCAGATCTTTGACTTCT	1288
Db	2221	AAAAANCCCCCCCCCTTCTTTTNTNAAAAAAATTTTTCATTA	2162
QY	1289	TGTTGTTAATGTTTATTTTGTAAAAATATAATTAATTAATTAATGAATGT	1348
Db	2161	AAGAAAGGAAATGTTTTTTTTTAAAAAAAGGCGGGGGGGTTT	2102
QY	1349	TTACAGCAACTCTCCCTAAAAAAAGGCGGT	1396
Db	2101	TTTTTTTTTCCCCCCCCCAAAAAAGAGGTGT	2054

Search completed: June 29, 2003, 04:55:58
Job time : 2489 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using bw model

Run on: June 28, 2003, 21:30:47 / Search time 246 Seconds
(without alignments)
12788.787 Million cell updates/sec

Title: US-09-989-919-15

Perfect score: 1397

Sequence: 1 ggtgctgcacgtctaccgga.....aaaaaaaaaagcggtc 1397

Scoring table: IDENTITY NUC

Gapop 10.0, Gapext 1.0

Searched: 2185239 seqs, 1125999159 residues

Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

N.Geneseq_101002.*
1: /SID22/gcgdata/geneq/geneq-emb1/NA1980.DAT.*
2: /SID22/gcgdata/geneq/geneq-emb1/NA1981.DAT.*
3: /SID22/gcgdata/geneq/geneq-emb1/NA1982.DAT.*
4: /SID22/gcgdata/geneq/geneq-emb1/NA1983.DAT.*
5: /SID22/gcgdata/geneq/geneq-emb1/NA1984.DAT.*
6: /SID22/gcgdata/geneq/geneq-emb1/NA1985.DAT.*
7: /SID22/gcgdata/geneq/geneq-emb1/NA1986.DAT.*
8: /SID22/gcgdata/geneq/geneq-emb1/NA1987.DAT.*
9: /SID22/gcgdata/geneq/geneq-emb1/NA1988.DAT.*
10: /SID22/gcgdata/geneq/geneq-emb1/NA1989.DAT.*
11: /SID22/gcgdata/geneq/geneq-emb1/NA1990.DAT.*
12: /SID22/gcgdata/geneq/geneq-emb1/NA1991.DAT.*
13: /SID22/gcgdata/geneq/geneq-emb1/NA1992.DAT.*
14: /SID22/gcgdata/geneq/geneq-emb1/NA1993.DAT.*
15: /SID22/gcgdata/geneq/geneq-emb1/NA1994.DAT.*
16: /SID22/gcgdata/geneq/geneq-emb1/NA1995.DAT.*
17: /SID22/gcgdata/geneq/geneq-emb1/NA1996.DAT.*
18: /SID22/gcgdata/geneq/geneq-emb1/NA1997.DAT.*
19: /SID22/gcgdata/geneq/geneq-emb1/NA1998.DAT.*
20: /SID22/gcgdata/geneq/geneq-emb1/NA1999.DAT.*
21: /SID22/gcgdata/geneq/geneq-emb1/NA2000.DAT.*
22: /SID22/gcgdata/geneq/geneq-emb1/NA2001A.DAT.*
23: /SID22/gcgdata/geneq/geneq-emb1/NA2001B.DAT.*
24: /SID22/gcgdata/geneq/geneq-emb1/NA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1284.8	92.0	9839	AA199052	Human excretory re
2	1284.8	92.0	9839	AA163402	Human kidney relat
3	952.4	68.2	1846	AA198172	Human EST-derived
4	940.4	67.3	2461	AA144725	Novel protein kina
5	420.8	30.1	4273	ABN94640	Gene #1138 used to
6	235.8	16.9	1886	AA144726	Novel protein kina
7	192.8	13.8	1837	AA178838	DNA encoding novel
8	149	10.7	150	AA186413	Human single nucle
9	149	10.7	150	AA186414	Human single nucle

C 10	59	4.2	429	23	ABV05669	Human prostate exp
C 11	54.6	3.9	370	23	AA178836	DNA encoding novel
C 12	51	3.7	400	23	ABV14838	Human prostate exp
C 13	51	3.7	436	23	ABV38926	Human prostate exp
C 14	51	3.7	436	23	ABV44714	Human prostate exp
C 15	51	3.7	83391	24	ABQ67093	Human angiotensin
C 16	50.6	3.6	425	22	AA160450	Human cancer agent
C 17	50.6	3.6	16688	24	ABL33320	Human immune syste
C 18	50.2	3.6	332	23	ABV56463	Human prostate exp
C 19	50.2	3.6	411	22	AA188595	Human polynucleoti
C 20	49.4	3.5	3795	22	AA191702	Mouse fatty acid e
C 21	48.8	3.5	309	23	ABV44994	Human prostate exp
C 22	48.8	3.5	556	23	ABV40063	Human prostate exp
C 23	48.8	3.5	556	23	ABV40163	Human prostate exp
C 24	48.8	3.5	556	23	ABV42105	Human prostate exp
C 25	48.8	3.5	556	23	ABV43601	Human prostate exp
C 26	48.6	3.5	389	22	AA120088	Human breast cancer
C 27	48.4	3.5	464	23	ABV03810	Human prostate exp
C 28	48.4	3.5	494	23	ABV10021	Human prostate exp
C 29	48.2	3.5	436	22	AA160532	Human cancer agent
C 30	48.2	3.5	8260	24	ABK28341	DNA transcription
C 31	48	3.4	337	22	AA115578	Human breast cancer
C 32	48	3.4	535	23	ABV03766	Human prostate exp
C 33	48	3.4	2040	22	AA107685	Human secreted pro
C 34	48	3.4	2040	22	AA104908	Human secreted pro
C 35	47.8	3.4	435	23	ABV60921	Human prostate exp
C 36	47.8	3.4	626	23	ABV60941	Human prostate exp
C 37	47.8	3.4	6071	24	AB192315	Chemically treated
C 38	47.8	3.4	6071	24	ABD22316	Chemically treated
C 39	47.6	3.4	457	22	AA188528	Human polynucleoti
C 40	47.4	3.4	165	22	AA184674	Human polynucleoti
C 41	47.4	3.4	300	22	AAH70049	Human cervical can
C 42	47.2	3.4	556	23	ABV40063	Human prostate exp
C 43	47.2	3.4	556	23	ABV40163	Human prostate exp
C 44	47.2	3.4	556	23	ABV42105	Human prostate exp
C 45	47.2	3.4	556	23	ABV43601	Human prostate exp

ALIGNMENTS

RESULT 1	AA199052	standard; DNA; 9839 BP.
ID	AA199052	standard; DNA; 9839 BP.
AC	AA199052;	
XX		
DT	07-JAN-2002	(first entry)
XX		
DE	Human excretory related polynucleotide seq	ID NO 816.
XX		
KW	Human; noctropic; neuroprotective; cyrostatic; dermatological; virucide;	
KW	Immunosuppressive; anti-HIV; antibacterial; vulnary;	
KW	anti-parasitic; antitubercular; antileishmanial; antileishmanial; cancer;	
KW	antitubercular; hepatotropic; cerebroprotective; antineoplastic;	
KW	antitubercular; antidiabetic; antileishmanial; antileishmanial; antileishmanial;	
KW	antiparasitic; candidant; immune disorder; cardiovascular disorder;	
KW	neurological disease; infection; nephrotropic; gene therapy; vaccine;	
KW	excretory system; ds.	
XX		
OS	Homo sapiens.	
XX		
PN	MO200155313-A2.	
XX		
PD	02-AUG-2001.	
XX		
PF	17-JAN-2001; 2001WO-US01323.	
XX		
PR	31-JAN-2000; 2000US-0179065.	
PR	04-FEB-2000; 2000US-0180628.	
PR	24-FEB-2000; 2000US-0184664.	
PR	02-MAR-2000; 2000US-0186350.	
PR	16-MAR-2000; 2000US-0189874.	

PR 17-MAR-2000; 2000US-0190076-
PR 18-APR-2000; 2000US-0198123-
PR 19-MAY-2000; 2000US-0205515-
PR 07-JUN-2000; 2000US-0209467-
PR 28-JUN-2000; 2000US-0214886-
PR 30-JUN-2000; 2000US-0215135-
PR 07-JUL-2000; 2000US-0216647-
PR 07-JUL-2000; 2000US-0216880-
PR 11-JUL-2000; 2000US-0217487-
PR 11-JUL-2000; 2000US-0217486-
PR 14-JUL-2000; 2000US-0218290-
PR 26-JUL-2000; 2000US-0220963-
PR 14-AUG-2000; 2000US-0224518-
PR 14-AUG-2000; 2000US-0224519-
PR 14-AUG-2000; 2000US-0225213-
PR 14-AUG-2000; 2000US-0225214-
PR 14-AUG-2000; 2000US-0225266-
PR 14-AUG-2000; 2000US-0225267-
PR 14-AUG-2000; 2000US-0225268-
PR 14-AUG-2000; 2000US-0225270-
PR 14-AUG-2000; 2000US-0225447-
PR 14-AUG-2000; 2000US-0225757-
PR 14-AUG-2000; 2000US-0225757-
PR 14-AUG-2000; 2000US-0225758-
PR 14-AUG-2000; 2000US-0225759-
PR 18-AUG-2000; 2000US-0226279-
PR 22-AUG-2000; 2000US-0226681-
PR 22-AUG-2000; 2000US-0226868-
PR 22-AUG-2000; 2000US-0227182-
PR 23-AUG-2000; 2000US-0227009-
PR 30-AUG-2000; 2000US-0228924-
PR 01-SEP-2000; 2000US-0229287-
PR 01-SEP-2000; 2000US-0229343-
PR 01-SEP-2000; 2000US-0229344-
PR 01-SEP-2000; 2000US-0229345-
PR 05-SEP-2000; 2000US-0229509-
PR 05-SEP-2000; 2000US-0229513-
PR 06-SEP-2000; 2000US-0230437-
PR 06-SEP-2000; 2000US-0230438-
PR 08-SEP-2000; 2000US-0231242-
PR 08-SEP-2000; 2000US-0231243-
PR 08-SEP-2000; 2000US-0231244-
PR 08-SEP-2000; 2000US-0231413-
PR 08-SEP-2000; 2000US-0231414-
PR 08-SEP-2000; 2000US-0232080-
PR 08-SEP-2000; 2000US-0232081-
PR 12-SEP-2000; 2000US-0231968-
PR 14-SEP-2000; 2000US-0232337-
PR 14-SEP-2000; 2000US-0232338-
PR 14-SEP-2000; 2000US-0232399-
PR 14-SEP-2000; 2000US-0232400-
PR 14-SEP-2000; 2000US-0233401-
PR 14-SEP-2000; 2000US-0233463-
PR 14-SEP-2000; 2000US-0233464-
PR 14-SEP-2000; 2000US-0233465-
PR 21-SEP-2000; 2000US-0234223-
PR 21-SEP-2000; 2000US-0234274-
PR 25-SEP-2000; 2000US-0234997-
PR 25-SEP-2000; 2000US-0234998-
PR 26-SEP-2000; 2000US-0235484-
PR 27-SEP-2000; 2000US-0235634-
PR 27-SEP-2000; 2000US-0235636-
PR 29-SEP-2000; 2000US-0236327-
PR 29-SEP-2000; 2000US-0236327-
PR 29-SEP-2000; 2000US-0236367-
PR 29-SEP-2000; 2000US-0236368-
PR 29-SEP-2000; 2000US-0236369-
PR 29-SEP-2000; 2000US-0236370-
PR 02-OCT-2000; 2000US-0237037-
PR 02-OCT-2000; 2000US-0237038-
PR 02-OCT-2000; 2000US-0237039-
PR 02-OCT-2000; 2000US-0237040-
PR 13-OCT-2000; 2000US-0239935-

PR 13-OCT-2000; 2000US-0239937-
PR 20-OCT-2000; 2000US-0240960-
PR 20-OCT-2000; 2000US-0241221-
PR 20-OCT-2000; 2000US-0241785-
PR 20-OCT-2000; 2000US-0241786-
PR 20-OCT-2000; 2000US-0241787-
PR 20-OCT-2000; 2000US-0241808-
PR 20-OCT-2000; 2000US-0241809-
PR 20-OCT-2000; 2000US-0241826-
PR 01-NOV-2000; 2000US-0244617-
PR 08-NOV-2000; 2000US-0246474-
PR 08-NOV-2000; 2000US-0246475-
PR 08-NOV-2000; 2000US-0246476-
PR 08-NOV-2000; 2000US-0246477-
PR 08-NOV-2000; 2000US-0246478-
PR 08-NOV-2000; 2000US-0246523-
PR 08-NOV-2000; 2000US-0246524-
PR 08-NOV-2000; 2000US-0246525-
PR 08-NOV-2000; 2000US-0246526-
PR 08-NOV-2000; 2000US-0246527-
PR 08-NOV-2000; 2000US-0246528-
PR 08-NOV-2000; 2000US-0246532-
PR 08-NOV-2000; 2000US-0246609-
PR 08-NOV-2000; 2000US-0246610-
PR 08-NOV-2000; 2000US-0246611-
PR 08-NOV-2000; 2000US-0246613-
PR 17-NOV-2000; 2000US-0249207-
PR 17-NOV-2000; 2000US-0249208-
PR 17-NOV-2000; 2000US-0249209-
PR 17-NOV-2000; 2000US-0249210-
PR 17-NOV-2000; 2000US-0249211-
PR 17-NOV-2000; 2000US-0249212-
PR 17-NOV-2000; 2000US-0249213-
PR 17-NOV-2000; 2000US-0249214-
PR 17-NOV-2000; 2000US-0249215-
PR 17-NOV-2000; 2000US-0249216-
PR 17-NOV-2000; 2000US-0249217-
PR 17-NOV-2000; 2000US-0249218-
PR 17-NOV-2000; 2000US-0249244-
PR 17-NOV-2000; 2000US-0249245-
PR 17-NOV-2000; 2000US-0249264-
PR 17-NOV-2000; 2000US-0249265-
PR 17-NOV-2000; 2000US-0249297-
PR 17-NOV-2000; 2000US-0249299-
PR 17-NOV-2000; 2000US-0249300-
PR 01-DEC-2000; 2000US-0250160-
PR 01-DEC-2000; 2000US-0250391-
PR 05-DEC-2000; 2000US-0251030-
PR 05-DEC-2000; 2000US-0251988-
PR 05-DEC-2000; 2000US-0256719-
PR 06-DEC-2000; 2000US-0251479-
PR 08-DEC-2000; 2000US-0251856-
PR 08-DEC-2000; 2000US-0251868-
PR 08-DEC-2000; 2000US-0251869-
PR 08-DEC-2000; 2000US-0251989-
PR 08-DEC-2000; 2000US-0251990-
PR 11-DEC-2000; 2000US-0256497-
PR 05-JAN-2001; 2001US-0259678-
XX
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Rosen CA, Barash SC, Ruben SM;
XX WPI; 2001-465569/50.
XX
XX Isolated nucleic acid molecule encoding excretory system antigen is
XX used in preventing, treating or ameliorating a medical condition -
PS Example 2; SEQ ID NO 816; 574pp + Sequence Listing; English.
XX
CC The invention relates to novel excretory system related human
CC polynucleotides (AA198567-AA199503) and the encoded proteins
CC (AA199594-AA199913) useful for preventing, treating or ameliorating

CC medical conditions e.g. by protein or gene therapy, especially
 CC disorders related to the excretory system. The genes are isolated
 CC from a range of human tissues disclosed in the specification. The
 CC nucleic acids, proteins, antibodies and (ant)agonists are useful in
 CC the diagnosis, treatment and prevention of: (a) cancer, e.g. breast
 CC and ovarian cancer and other cancers of the adrenal gland, bone, bone
 CC marrow, breast, gastrointestinal tract, liver, lung, or urogenital;
 CC (b) immune disorders e.g. Addison's disease, allergies, autoimmune
 CC haemolytic anaemia, autoimmune thyroiditis, diabetes mellitus, Crohn's
 CC disease, multiple sclerosis, rheumatoid arthritis and ulcerative
 CC colitis; (c) cardiovascular disorders such as myocardial ischaemia;
 CC (d) wound healing; (e) neurological diseases e.g. cerebral anoxia and
 CC epilepsy; and (f) infectious diseases such as viral, bacterial, fungal
 CC and parasitic infections.
 CC Note: The sequence data for this patent did not form part of the
 CC printed specification, but was obtained in electronic format directly
 CC from WHO at ftp.wipo.int/pub/published_pct_sequences.

XX Sequence 9839 BP; 2292 A; 2702 C; 2654 G; 2191 T; 0 other;

Query Match 92.0%; Score 1284.8; DB 22; Length 9839;

Best Local Similarity 99.7%; Pred. No. 1.1e-289; Mismatches 2; Gaps 2;

Matches 1308; Conservative 0; Indels 2; Gaps 2;

QY 59 ACCGAGTACCACTGTATCCAGACAGACCCATCCCGAGAGACTACCGCTGCGCCA 118
 DB 8530 ACAGAGTACCACTGTATCCAGACAGACCCATCCCGAGAGACTACCGCTGCGCCA 8589

QY 119 TCTTACCAACAGCGGAGCTGCTCTTCACTGTTCACTGCTGAGCTGTGATGTC 178
 DB 8590 TCTTACCAACAGCGGAGCTGCTCTTCACTGTTCACTGCTGAGCTGTGATGTC 8649

QY 179 TGTGAGAGCCATGCCAGTGTGGGCTTGTGTGTACCAACAGACCACTGTGAGCTG 238
 DB 8650 TGTGAGAGCCATGCCAGTGTGGGCTTGTGTGTACCAACAGACCACTGTGAGCTG 8709

QY 239 GAGCCAGTGGAGAGAGCCCTTCCAGAGAGATGGAGAGACCTCTGAGAGTTGATGAT 298
 DB 8710 GAGCCAGTGGAGAGAGCCCTTCCAGAGAGATGGAGAGACCTCTGAGAGTTGATGAT 8769

QY 299 AGTGATCCCTCCATCGGAGTCAAGAGGGGCTGTGAGTGTGAGAGAGTATACGTGT 358
 DB 8770 AGTGATCCCTCCATCGGAGTCAAGAGGGGCTGTGAGTGTGAGAGAGTATACGTGT 8829

QY 359 CTTCAAGGAGCAATTAATTTGGAGAGATGTCTTCCCGAGAGAGAAATCCAGCC 418
 DB 8830 CTTCAAGGAGCAATTAATTTGGAGAGATGTCTTCCCGAGAGAGAAATCCAGCC 8889

QY 419 TGTTACCTCTCACTGCTGCCCCCGAGTGGAGAGTGTCTTTTCAAGACTGGATGAG 478
 DB 8890 TGTTACCTCTCACTGCTGCCCCCGAGTGGAGAGTGTCTTTTCAAGACTGGATGAG 8949

QY 479 CCAAGTGTCTCTGATTCACCAAGACACATATGTGAGAGCTCTGCTGACTATCTG 538
 DB 8950 CCAAGTGTCTCTGATTCACCAAGACACATATGTGAGAGCTCTGCTGACTATCTG 9008

QY 539 AGGGCTCGGTGACCAAGTCAATCTCTGAGAGCTGGGCTTCCCTGTGAGAGAGTGC 598
 DB 9009 AGGGCTCGGTGACCAAGTCAATCTCTGAGAGCTGGGCTTCCCTGTGAGAGAGTGC 9068

QY 599 TTGCACTGGAGAGACTGATGTCACTGGAGAGCCCTGAGAGAGAACTTCCAG 658
 DB 9069 TTGCACTGGAGAGACTGATGTCACTGGAGAGCCCTGAGAGAGAACTTCCAG 9128

QY 659 CAGACAGATGTGACCAAGACAAAGTCAATTAATTTGAAATGTGATTTACC 718
 DB 9129 CAGACAGATGTGACCAAGACAAAGTCAATTAATTTGAAATGTGATTTACC 9188

QY 719 AGCTTGTGATGTGAGAGTGTCTCTGATTCACCAAGATATGTGGGATATGCTCTTC 778
 DB 9189 AGCTTGTGATGTGAGAGTGTCTCTGATTCACCAAGATATGTGGGATATGCTCTTC 9248

QY 779 CAACCTGTGGGCTGTGAGAGAGCTAGCTATCTTCCCACTGGGGGCTGTCCCTCC 838

DB 9249 CAACCTGTGGGCTGTGAGAGAGCTAGCTAGTATCCCACT-GGGGCTGTCCCTCC 9307

QY 839 CTGGAGAGGTTCCGTGGAGAGCCCTCACTGTGTTCAATAGTGTAGAAATGACTAA 898

DB 9308 CTGGAGAGGTTCCGTGGAGAGCCCTCACTGTGTTCAATAGTGTAGAAATGACTAA 9367

QY 899 GCCCTGTGCTGCTCTGCTGACATGACAGAGAGCGGTGGGGGCTGCGTGGAGCAATC 958

DB 9368 GCCCTGTGCTGCTCTGCTGACATGACAGAGAGCGGTGGGGGCTGCGTGGAGCAATC 9427

QY 959 CATGTGAGTGTCTCTGAGCTTGTGAGACAGAGACTTGGCGGGAGATGCTCCAG 1018

DB 9428 CATGTGAGTGTCTCTGAGCTTGTGAGACAGAGACTTGGCGGGAGATGCTCCAG 9487

QY 1019 GATGTGGGTGATTTCTTACTGAGAGGCTATCTGTGACTCCCGACAGGGACATCC 1078

DB 9488 GATGTGGGTGATTTCTTACTGAGAGGCTATCTGTGACTCCCGACAGGGACATCC 9547

QY 1079 AGGCGAGCCAGGGGTGAGAGAGTGCACACCTCAGCATGAGCCAAAGACTGGGCTC 1138

DB 9548 AGGCGAGCCAGGGGTGAGAGAGTGCACACCTCAGCATGAGCCAAAGACTGGGCTC 9607

QY 1139 AGGAGCAGTGTGTTTGTGAGCCAGAGCTTGGGCGGGGCTGGGCTTTCTCC 1198

DB 9608 AGGAGCAGTGTGTTTGTGAGCCAGAGCTTGGGCGGGGCTGGGCTTTCTCC 9667

QY 1199 CTGATTTGCTTCAATGAAAGCTCAAGAGCAGCAAAACAGGCTTTCCCTCTCGA 1258

DB 9668 CTGATTTGCTTCAATGAAAGCTCAAGAGCAGCAAAACAGGCTTTCCCTCTCGA 9727

QY 1259 GTTGAATATCCAGAACTTTTGTACTCTGTGTGTTAATGTTATTTTGTAAATA 1318

DB 9728 GTTGAATATCCAGAACTTTTGTACTCTGTGTGTTAATGTTATTTTGTAAATA 9787

QY 1319 ATAAATTAATTAATTAATTAATTAATGATGTTTCAAGCAAACTTTCCCTAA 1370

DB 9788 ATAAATTAATTAATTAATTAATGATGTTTCAAGCAAACTTTCCCTAA 9839

RESULT 2
 AA163402
 ID AA163402 standard; DNA; 9839 BP.
 XX
 AC AA163402;
 XX
 DT 22-OCT-2001 (first entry)
 XX
 DE Human kidney related polynucleotide seq ID NO 717.
 XX
 KW Human; kidney antigen; immunosuppressive; antiarthritic; antirheumatic;
 KW antiproliferative; cytostatic; cardiac; vasodilator; cerebroprotective;
 KW neurotropic; neuroprotective; antibacterial; virulence; fungicide;
 KW ophthalmological; antiallergic; hepatocytic; antidiabetic;
 KW antiinflammatory; antitumor; antileukemic; antiparasitic;
 KW Gene therapy; cancer; immune disorder; cardiovascular disorder;
 KW neurological disease; infection; ds.
 XX
 OS Homo sapiens.
 XX
 PN MO200155323-A2.
 PD 02-AUG-2001.
 XX
 XX 17-JAN-2001; 2001WO-US01343.
 PF 31-JAN-2000; 2000US-0179065.
 PR 04-FEB-2000; 2000US-0180628.
 PR 24-FEB-2000; 2000US-0184664.
 PR 02-MAR-2000; 2000US-0186350.
 PR 16-MAR-2000; 2000US-0189874.
 PR 17-MAR-2000; 2000US-0190076.
 PR 18-APR-2000; 2000US-0198123.

PR 19-MAY-2000; 2000US-0205515.
 PR 07-JUN-2000; 2000US-0209467.
 PR 28-JUN-2000; 2000US-0214886.
 PR 30-JUN-2000; 2000US-0215135.
 PR 07-JUL-2000; 2000US-0216647.
 PR 07-JUL-2000; 2000US-0216880.
 PR 11-JUL-2000; 2000US-0217487.
 PR 11-JUL-2000; 2000US-0217496.
 PR 14-JUL-2000; 2000US-0218220.
 PR 26-JUL-2000; 2000US-0220963.
 PR 26-JUL-2000; 2000US-0220964.
 PR 14-AUG-2000; 2000US-0224518.
 PR 14-AUG-2000; 2000US-0224519.
 PR 14-AUG-2000; 2000US-0225213.
 PR 14-AUG-2000; 2000US-0225214.
 PR 14-AUG-2000; 2000US-0225266.
 PR 14-AUG-2000; 2000US-0225267.
 PR 14-AUG-2000; 2000US-0225270.
 PR 14-AUG-2000; 2000US-0225447.
 PR 14-AUG-2000; 2000US-0225757.
 PR 14-AUG-2000; 2000US-0225758.
 PR 14-AUG-2000; 2000US-0225759.
 PR 18-AUG-2000; 2000US-0226279.
 PR 22-AUG-2000; 2000US-0226681.
 PR 22-AUG-2000; 2000US-0226681.
 PR 22-AUG-2000; 2000US-0227182.
 PR 23-AUG-2000; 2000US-0227009.
 PR 30-AUG-2000; 2000US-0228924.
 PR 01-SEP-2000; 2000US-0229287.
 PR 01-SEP-2000; 2000US-0229343.
 PR 01-SEP-2000; 2000US-0229344.
 PR 01-SEP-2000; 2000US-0229345.
 PR 05-SEP-2000; 2000US-0229509.
 PR 05-SEP-2000; 2000US-0229513.
 PR 06-SEP-2000; 2000US-0230438.
 PR 08-SEP-2000; 2000US-0231242.
 PR 08-SEP-2000; 2000US-0231243.
 PR 08-SEP-2000; 2000US-0231244.
 PR 08-SEP-2000; 2000US-0231413.
 PR 08-SEP-2000; 2000US-0232080.
 PR 08-SEP-2000; 2000US-0232081.
 PR 12-SEP-2000; 2000US-0231968.
 PR 14-SEP-2000; 2000US-0232387.
 PR 14-SEP-2000; 2000US-0232388.
 PR 14-SEP-2000; 2000US-0232389.
 PR 14-SEP-2000; 2000US-0232400.
 PR 14-SEP-2000; 2000US-0232401.
 PR 14-SEP-2000; 2000US-0233063.
 PR 14-SEP-2000; 2000US-0233064.
 PR 14-SEP-2000; 2000US-0233065.
 PR 21-SEP-2000; 2000US-0234223.
 PR 21-SEP-2000; 2000US-0234274.
 PR 25-SEP-2000; 2000US-0234997.
 PR 25-SEP-2000; 2000US-0234998.
 PR 26-SEP-2000; 2000US-0235484.
 PR 27-SEP-2000; 2000US-0235834.
 PR 27-SEP-2000; 2000US-0235836.
 PR 29-SEP-2000; 2000US-0236337.
 PR 29-SEP-2000; 2000US-0236367.
 PR 29-SEP-2000; 2000US-0236368.
 PR 29-SEP-2000; 2000US-0236369.
 PR 29-SEP-2000; 2000US-0236370.
 PR 02-OCT-2000; 2000US-0237037.
 PR 02-OCT-2000; 2000US-0237038.
 PR 02-OCT-2000; 2000US-0237039.
 PR 02-OCT-2000; 2000US-0237040.
 PR 13-OCT-2000; 2000US-0239935.
 PR 13-OCT-2000; 2000US-0239937.
 PR 20-OCT-2000; 2000US-0240960.

PR 20-OCT-2000; 2000US-0241221.
 PR 20-OCT-2000; 2000US-0241785.
 PR 20-OCT-2000; 2000US-0241786.
 PR 20-OCT-2000; 2000US-0241787.
 PR 20-OCT-2000; 2000US-0241808.
 PR 20-OCT-2000; 2000US-0241809.
 PR 20-OCT-2000; 2000US-0241826.
 PR 01-NOV-2000; 2000US-0244617.
 PR 08-NOV-2000; 2000US-0246474.
 PR 08-NOV-2000; 2000US-0246475.
 PR 08-NOV-2000; 2000US-0246476.
 PR 08-NOV-2000; 2000US-0246477.
 PR 08-NOV-2000; 2000US-0246478.
 PR 08-NOV-2000; 2000US-0246523.
 PR 08-NOV-2000; 2000US-0246524.
 PR 08-NOV-2000; 2000US-0246525.
 PR 08-NOV-2000; 2000US-0246526.
 PR 08-NOV-2000; 2000US-0246527.
 PR 08-NOV-2000; 2000US-0246528.
 PR 08-NOV-2000; 2000US-0246532.
 PR 08-NOV-2000; 2000US-0246609.
 PR 08-NOV-2000; 2000US-0246610.
 PR 08-NOV-2000; 2000US-0246611.
 PR 17-NOV-2000; 2000US-0246613.
 PR 17-NOV-2000; 2000US-0249207.
 PR 17-NOV-2000; 2000US-0249208.
 PR 17-NOV-2000; 2000US-0249209.
 PR 17-NOV-2000; 2000US-0249210.
 PR 17-NOV-2000; 2000US-0249211.
 PR 17-NOV-2000; 2000US-0249212.
 PR 17-NOV-2000; 2000US-0249213.
 PR 17-NOV-2000; 2000US-0249214.
 PR 17-NOV-2000; 2000US-0249215.
 PR 17-NOV-2000; 2000US-0249216.
 PR 17-NOV-2000; 2000US-0249217.
 PR 17-NOV-2000; 2000US-0249218.
 PR 17-NOV-2000; 2000US-0249219.
 PR 17-NOV-2000; 2000US-0249297.
 PR 17-NOV-2000; 2000US-0249299.
 PR 01-DEC-2000; 2000US-0250160.
 PR 01-DEC-2000; 2000US-0250391.
 PR 05-DEC-2000; 2000US-0251030.
 PR 05-DEC-2000; 2000US-0251988.
 PR 06-DEC-2000; 2000US-0251479.
 PR 08-DEC-2000; 2000US-0251856.
 PR 08-DEC-2000; 2000US-0251868.
 PR 08-DEC-2000; 2000US-0251869.
 PR 08-DEC-2000; 2000US-0251989.
 PR 11-DEC-2000; 2000US-0254097.
 PR 05-JAN-2001; 2001US-02559678.
 XX (HUMA-) HUMAN GENOME SCI INC.
 XX
 XX
 PI Rosen CA, Barash SC, Ruben SM,
 PI
 XX
 DR WPI; 2001-488784/53.
 XX
 PT New isolated nucleic acids and polypeptides, useful for diagnosing,
 PT treating and/or preventing human diseases and disorders -
 XX
 PS Disclosure; SEQ ID NO 717; 564pp + Sequence Listing; English.
 XX
 CC The invention relates to novel kidney related polynucleotides
 CC (AA162971-AA163793) and the encoded polypeptides (AA162417-AA162691)
 CC collectively known as kidney antigens and the use of such kidney antigens
 CC for detecting disorders of the kidney, especially kidney cancer and
 CC kidney cancer metastases. The polynucleotides and proteins are also

CC useful for preventing, treating or ameliorating medical conditions
 CC e.g. by protein or gene therapy. The genes are isolated from a range
 CC of human tissues disclosed in the specification. The nucleic acids,
 CC proteins, antibodies and (ant)agonists are useful in the diagnosis,
 CC treatment and prevention of: (a) cancer, e.g. breast and ovarian cancer,
 CC and other cancers of the adrenal gland, bone, bone marrow, breast,
 CC gastrointestinal tract, liver, lung, or urogenital; (b) immune disorders
 CC e.g. Addison's disease, allergies, autoimmune haemolytic anaemia,
 CC autoimmune thyroiditis, diabetes mellitus, Crohn's disease, multiple
 CC sclerosis, rheumatoid arthritis and ulcerative colitis;
 CC (c) cardiovascular disorders such as myocardial ischaemia; (d) wound
 CC healing; (e) neurological diseases e.g. cerebral anoxia and epilepsy;
 CC and (f) infectious diseases such as viral, bacterial, fungal and
 CC parasitic infections.
 CC Note: The sequence data for this patent did not form part of the
 CC printed specification, but was obtained in electronic format directly
 CC from WHO at ftp.wipo.int/pub/published_pct_sequences.

XX Sequence 9839 BP; 2292 A; 2702 C; 2654 G; 2191 T; 0 other;

Query Match 92.0%; Score 1284.8; DB 22; Length 9839;
 Best Local Similarity 99.7%; Pred. No. 1.1e-289;
 Matches 1308; Conservative 0; Mismatches 2; Indels 2; Gaps 2;

QY 59 ACCGAGTACCGATGTATCCGACAGACACCATCCCGAGAGACTACCGCTGCGCA 118
 DB 8530 ACAGATACCAAGTATCCGACAGACACCATCCCGAGAGACTACCGCTGCGCA 8589
 QY 119 TCTTACCAACGAGGAGCTGCTCTTCACTGTTCACTGCTGAGGCTGATGTC 178
 DB 8590 TCTTACCAACGAGGAGCTGCTCTTCACTGTTCACTGCTGAGGCTGATGTC 8649
 QY 179 TGTGAGAGCATCCGAGTGTGGGGCTTGTGTGACCAACAGACCACTGGAAGGT 238
 DB 8650 TGTGAGAGCATCCGAGTGTGGGGCTTGTGTGACCAACAGACCACTGGAAGGT 8709
 QY 239 GAGCCAGTGGAGAGACCTTCCAGAGAGATGGAGACCTCTCTGAGGTTATGAT 298
 DB 8710 GAGCCAGTGGAGAGACCTTCCAGAGAGATGGAGACCTCTCTGAGGTTATGAT 8769
 QY 299 AGTGATCCCCCATCGGAAGTCAAGGGGGTGTGAGGTGATGAGAGAGTATACGT 358
 DB 8770 AGTGATCCCCCATCGGAAGTCAAGGGGGTGTGAGGTGATGAGAGAGTATACGT 8829
 QY 359 CTTCAAGGAGCAATTAAGGAGATGTCTTCCAGAGAGAGAGAGAGATCCAGCCC 418
 DB 8830 CTTCAAGGAGCAATTAAGGAGATGTCTTCCAGAGAGAGAGAGAGATCCAGCCC 8889
 QY 419 TGTTACCTCTACCTCTGCCCCCAGAGTGGAGAGTGTCTTTTCAAGACTGATGAG 478
 DB 8890 TGTTACCTCTACCTCTGCCCCCAGAGTGGAGAGTGTCTTTTCAAGACTGATGAG 8949
 QY 479 CCAAGTGTCTCTGATCCCAACAGACCAATATGTAAGGCTCTGCTGATCTTCTG 538
 DB 8950 CCAAGTGTCTCTGATCCCAACAGACCAATATGTAAGGCTCTGCTGATCTTCTG 9008
 QY 539 AGGGCTCGGCTGACACGCTGACTATCTTCAAGAGCTGGGCTTCTGCTGAGGAGT 598
 DB 9009 AGGGCTCGGCTGACACGCTGACTATCTTCAAGAGCTGGGCTTCTGCTGAGGAGT 9068
 QY 599 TTGCACTGGCAGACTGATGTCACTGGGAGACCCCTGAGACAAAGCTATCCAG 658
 DB 9069 TTGCACTGGCAGACTGATGTCACTGGGAGACCCCTGAGACAAAGCTATCCAG 9128
 QY 659 CAGACAGATGTGACAGAGCAAAAGTGCATTAATGCAATGTTAAATGTGATTAC 718
 DB 9129 CAGACAGATGTGACAGAGCAAAAGTGCATTAATGCAATGTTAAATGTGATTAC 9188
 QY 719 AGCTAGCTATGGGAGCTGCTGCTCTAGTCCAGAGATATGAGGCTATGACCTCTC 778
 DB 9189 AGCTAGCTATGGGAGCTGCTGCTCTAGTCCAGAGATATGAGGCTATGACCTCTC 9248
 QY 779 CAACCTGTGGCTGTAGCAAGCTCAGGCTAGTCTCCCACTGGGGGCTGTGCCCTCC 838

DB 9249 CAACCTGTGGCTGTAGCAAGCTCAGGCTAGTCTCCCACT-GGGGCTGTGCCCTCC 9307
 QY 839 CTGGGAGGTTCCGAGGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 898
 DB 9308 CTGGGAGGTTCCGAGGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 9367
 QY 899 GCCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 958
 DB 9368 GCCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 9427
 QY 959 CATGCTGAGTGTCTCTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAG 1018
 DB 9428 CATGCTGAGTGTCTCTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAG 9487
 QY 1019 GATGTGGGATTTCTGATCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1078
 DB 9488 GATGTGGGATTTCTGATCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 9547
 QY 1079 AGGCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1138
 DB 9548 AGGCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 9607
 QY 1139 AGGAGCAGTGTGTTGAGCCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1198
 DB 9608 AGGAGCAGTGTGTTGAGCCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 9667
 QY 1199 CTCTTGTCTTCAATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1258
 DB 9668 CTCTTGTCTTCAATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 9727
 QY 1259 GTTGAATATCCAGATCTTGTGATCTTGTGATCTTGTGATCTTGTGATCTTGTG 1318
 DB 9728 GTTGAATATCCAGATCTTGTGATCTTGTGATCTTGTGATCTTGTGATCTTGTG 9787
 QY 1319 ATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 1370
 DB 9788 ATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 9839

RESULT 3
 AAH98172
 ID AAH98172 standard; cDNA; 1846 BP.
 XX
 AC AAH98172;
 DT 12-OCT-2001 (first entry)
 DE
 XX
 DE Human EST-derived coding sequence SEQ ID NO: 29.
 XX
 XX Human; sheep; pig; cow; fruit fly; yeast; hamster; macaque; horse;
 KW tomato; monkey; dog; sea urchin; expressed sequence tag; EST;
 KW diagnostics; forensic test; gene mapping; genetic disorder;
 KW biodiversity; gene therapy; nutrition; ss.
 OS Homo sapiens.
 PN WO200154477-A2.
 XX
 PD 02-AUG-2001.
 XX
 PF 25-JAN-2001; 2001WO-US02687.
 XX
 XX 25-JAN-2000; 2000US-0491404.
 PR 17-JUL-2000; 2000US-0617746.
 PR 03-AUG-2000; 2000US-0631451.
 PR 15-SEP-2000; 2000US-0663870.
 XX
 PA (HYSE-) HYSEQ INC.
 XX
 XX Tang YT, Liu C, Zhou P, Qian XB, Wang Z, Chen R, Asundi V;
 PI Cao Y, Drmanac RA, Zhang J, Wehrman T;
 XX

DR MPI: 2001-476164/51.
 DR P-PSDB: AAM23513.
 XX Isolated polypeptide for treatment of diseases, diagnostics, raising
 PT antibodies and research use -
 XX
 XX Claim 1; Page 202-203; 1275bp; English.
 CC The present invention provides the protein and coding sequences of novel
 CC proteins from a variety of organisms, including human, dog, cat, horse,
 CC cow, pig, hamster, monkey, macaque, yeast, bacteria, fruit fly, sea
 CC urchin and tomato. These were derived from expressed sequence tags (ESTs)
 CC from the organism of interest. They can be used in diagnostics,
 CC forensics, gene mapping, identification of mutations, to assess
 CC biodiversity and for nutritional purposes. The present sequence is a cDNA
 CC of the invention.
 XX
 XX Sequence 1846 BP; 397 A; 526 C; 538 G; 385 T; 0 other;
 SQ
 Query Match 68.2%; Score 952.4; DB 22; Length 1846;
 Best Local Similarity 84.9%; Pred. No. 2.3e-212;
 Matches 1171; Conservative 0; Mismatches 1; Indels 208; Gaps 1;

QY 1 GGTGCTGACCTGTACCGAGCGGCGAGATCTGCAGAACTCCACGCGAAGACAGCACTAC 60
 DB 675 GGTGCTGACCTGTACCGAGCGGCGAGATCTGCAGAACTCCACGCGAAGACAGCACTAC 734
 QY 61 CGAGTACAGTGTATCCCAACAGACACATCCCGGAGAAAGTACCTGCTGCGCATC 120
 DB 735 CGAGTACAGTGTATCCCAACAGACACATCCCGGAGAAAGTACCTGCTGCGCATC 794
 QY 121 CTACACACAGGAGAGCTGCTCTTCAAGTTCACCTGAGGCTGAGATGCTG 180
 DB 795 CTACACACAGGAGAGCTGCTCTTCAAGTTCACCTGAGGCTGAGATGCTG 854
 QY 181 TGAAGAGCATCCAGATGTCGGGCTTTGTGTACCAACAGACCACTGACAGGTGA 240
 DB 855 TGAAGAGCATCCAGATGTCGGGCTTTGTGTACCAACAGACCACTGACAGGTGA 907
 QY 241 GCCAGTGGAGAAAGCCCTTCCAGAGGAGATGGACGAGACTCTCTGAGGTTGATAGTAG 300
 DB 908 ----- 907
 QY 301 TGATCCCCCATCGAAGTCAAGGGGGTGTGATGATGAGAGAGATATAGTGTCT 360
 DB 908 ----- 907
 QY 361 TCAAGGAGTCAATTAAGGAGATGTCTTGTCCAGAAAGAAACATCCAGCCCTG 420
 DB 908 ----- 907
 QY 421 TTACCTCTACCTCTGCCCCCAGGTGGAGAGTGTCTTTTTCAGATCTGATGAGCC 480
 DB 908 ----- 946
 QY 481 AAGTGGCTGTATCCCAAGACCAATATGAGAGGCTGCTGCTGATCTGAG 540
 DB 947 AAGTGGCTGTATCCCAAGACCAATATGAGAGGCTGCTGCTGATCTGAG 1006
 QY 541 GGTCTGGCTGACAGCTGATCTATCTCAGAGCTGGGCTTGTCTGAGAGGAGTACTT 600
 DB 1007 GGTCTGGCTGACAGCTGATCTATCTCAGAGCTGGGCTTGTCTGAGAGGAGTACTT 1066
 QY 601 GCACTGGCAGACCTGATCTCACTTGGGAAACCTCTGAGACAAAGTAAATCCAGACA 660
 DB 1067 GCACTGGCAGACCTGATCTCACTTGGGAAACCTCTGAGACAAAGTAAATCCAGACA 1126
 QY 661 GACAGATGTACAGAGCAAAAGTGAATATGCAATATGTAATGATGTTACAG 720
 DB 1127 GACAGATGTACAGAGCAAAAGTGAATATGCAATATGTAATGATGTTACAG 1186
 QY 721 CCTAGATGTGAGCTGCTCTTATCTCAGAGATCATGGGGTATGACTGCTCTCA 780

DB 1187 CCTAGATGTGAGCTGCTCTTATCTCAGAGATCATGGGGTATGACTGCTCTCA 1246
 QY 781 ACCCTGGAGCTGTAGCAAGCTCAGGGTATGCTCCCACTGGGGCTGACCTCCCT 840
 DB 1247 ACCCTGGAGCTGTAGCAAGCTCAGGGTATGCTCCCACTGGGGCTGACCTCCCT 1306
 QY 841 GGGAGCGTTCCTGGGAGAGCCCATCATCTGTGTCATATGATGAGAAATGATAGTAAAG 900
 DB 1307 GGGAGCGTTCCTGGGAGAGCCCATCATCTGTGTCATATGATGAGAAATGATAGTAAAG 1366
 QY 901 CCTGCTGCTGCTGCTGCAATGCCACAGAGCGGCTGCTGCTGCTGCTGCTGCTGCTGCT 960
 DB 1367 CCTGCTGCTGCTGCTGCAATGCCACAGAGCGGCTGCTGCTGCTGCTGCTGCTGCTGCT 1426
 QY 961 TCGTGAAGTGTCTCTCAGCTTATGATGATGATGATGATGATGATGATGATGATGATGAT 1020
 DB 1427 TCGTGAAGTGTCTCTCAGCTTATGATGATGATGATGATGATGATGATGATGATGATGAT 1486
 QY 1021 TGTGGGTGATTCGTATCTGATCTGGGAGGCTATCTCTGACCTCCGACAGGGGACACTCCAG 1080
 DB 1487 TGTGGGTGATTCGTATCTGATCTGGGAGGCTATCTCTGACCTCCGACAGGGGACACTCCAG 1546
 QY 1081 GCCAGCCAGGGGCTCAGGGGAGAGGTGACACCTCAGATGACCAAGATCTGGGCTCAG 1140
 DB 1547 GCCAGCCAGGGGCTCAGGGGAGAGGTGACACCTCAGATGACCAAGATCTGGGCTCAG 1606
 QY 1141 GAGAGAGTGTGTTGACCTGAGACCTGAGGCGGGGCGGGGCGGGGCGGGGCGGGGCGGGG 1200
 DB 1607 GAGAGAGTGTGTTGACCTGAGACCTGAGGCGGGGCGGGGCGGGGCGGGGCGGGGCGGGG 1666
 QY 1201 CATTGCTTCAATGAAAGCTCAAAAGACCAAAACAGAGCTTCCCTCTCTGAGT 1260
 DB 1667 CATTGCTTCAATGAAAGCTCAAAAGACCAAAACAGAGCTTCCCTCTCTGAGT 1726
 QY 1261 TTGAATATCCAGATCTTTTGTATCTTGTGTTAAATGTTATTTTGTAAATAAT 1320
 DB 1727 TTGAATATCCAGATCTTTTGTATCTTGTGTTAAATGTTATTTTGTAAATAAT 1786
 QY 1321 AAAATTAATTTAGTAAATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1380
 DB 1787 AAAATTAATTTAGTAAATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1846

RESULT 4
 AAF44725
 ID AAF44725 standard; cDNA; 2461 BP.
 XX AAF44725;
 AC
 XX
 DT 27-MAR-2001 (first entry)
 XX
 XX Novel protein kinase cDNA, SEQ ID NO: 106.
 DE
 XX Human; mouse; protein kinase; antiarthritic; antisclerotic; osteopathetic;
 XX immunosuppressive; candidant; antiinflammatory; antiasthmatic;
 XX dermatological; antidiabetic; antifertility; gene therapy; vaccine;
 XX immune disorder; cardiovascular disease; neurodegenerative disease;
 XX cancer; autoimmune disorder; stroke; inflammatory bowel disease;
 XX inflammatory pelvic disease; multiple sclerosis; psoriasis; ss.
 OS Homo sapiens.
 XX
 XX MO200073469-A2.
 PN
 XX
 PD 07-DEC-2000.
 XX
 XX 26-MAY-2000; 2000WC-US14842.
 PF
 XX
 PR 28-MAY-1999; 99US-0136503.
 XX
 XX (SUGEN-) SUGEN INC.
 PA
 XX Plowman GD, Martinez R, Whyte D, Sudersanam S;

XX WPI, 2001-032161/04.
DR P-PSDB; AAB65697.

XX Nucleic acids encoding kinase polypeptides, useful for diagnosing and
PT treating immune-related diseases and disorders, cardiovascular disease,
PT neurodegenerative diseases and/or cancers -

XX Example 1; Fig 2; 310pp; English.

XX The present sequence encodes a novel protein kinase. The nucleic acids
CC and the protein kinases they encode may be used in the treatment and
CC diagnosis of diseases associated with inappropriate kinase expression
CC such as immune-related diseases and disorders, cardiovascular disease,
CC neurodegenerative diseases and/or cancers. The nucleic acids and
CC complementary sequences may also be used as DNA probes in diagnostic
CC assays. The kinase polypeptides may be used as antigens in the production
CC of antibodies of kinase expression and activity. Anti-kinase antibodies
CC and kinase antagonists may also be used to down regulate kinase
CC expression and activity. Diseases related to kinase expression and
CC activity include rheumatoid arthritis, atherosclerosis, autoimmune
CC disorders, complications of organ transplantation, myocardial infarction,
CC immune disorders, cardiomyopathies, strokes, renal failure,
CC oxidative stress related disorders, chronic inflammatory bowel disease,
CC chronic inflammatory pelvic disease, multiple sclerosis, asthma,
CC osteoarthritis, psoriasis, rhinitis, autoimmunity, diabetes, cancers and
CC reproductive disorders.

XX Sequence 2461 BP; 426 A; 771 C; 811 G; 453 T; 0 other;

XX Query Match 67.3%; Score 940.4; DB 22; Length 2461;

XX Best Local Similarity 84.7%; Pred. No. 1.6e-209;

XX Matches 1159; Conservative 0; Mismatches 1; Indels 208; Gaps 1;

QY 1 GGTGCTGACCTGATCCGAGCGGGGAGTATCTGCAAGATCCAGCGGACGACGATAC 60
DB 1302 GGTGCTGACCTGATCCGAGCGGGGAGTATCTGCAAGATCCAGCGGACGACGATAC 1361
QY 61 CGAGTACCACTGATATCCAGACAGACATCCCGAGAGAGACTACCGCTGCGCATC 120
DB 1362 CGAGTACCACTGATATCCAGACAGACATCCCGAGAGAGACTACCGCTGCGCATC 1421
QY 121 CTACACACAGGAGAGCTGCTCTTTCAGTTCATCACTGAGGCTGTGATGCTG 180
DB 1422 CTACACACAGGAGAGCTGCTCTTTCAGTTCATCACTGAGGCTGTGATGCTG 1481
QY 181 TGAGAGCCATGCCAGTGTGGGCTTTGTGTGTCACCAACCAACCACTGACAGGTGA 240
DB 1482 TGAGAGCCATGCCAGTGTGGGCTTTGTGTGTCACCAACCAACCACTGACAGGTGA 1534
QY 241 GCCAGTGGAGAGAGCCCTTCCAGAGGAGATGGACGACCTCTCGAGGTTGATAGTAG 300
DB 1535 ----- 1534
QY 301 TGATCCCCCATCGAAGTCAAGGGGGGTGCTGAGTGATGAGAGAGATATACGTGTCT 360
DB 1535 ----- 1534
QY 361 TCAAGGACATCAATTATGGAGATGTCTTGCTCCAGAAAGAAACATCCAGCCCTG 420
DB 1535 ----- 1534
QY 421 TTACTCTTCACTCTGCCCCCAGGTGGAGAGTGTCTTTTCAAGACTGATGGAGCC 480
DB 1535 ----- 1534
QY 481 AAGTGTCCCTGATCCCAACAGACCAATATGTAAGGCTGTGCTGATGATCTGAG 540
DB 1574 AAGTGTCCCTGATCCCAACAGACCAATATGTAAGGCTGTGCTGATGATCTGAG 1633
QY 541 GGCTCGGCTTACCAAGCTGATATCTTCAAGAGCTGGGCTTGCTGTGAGAGGATGACTT 600
DB 1634 GGCTCGGCTTACCAAGCTGATATCTTCAAGAGCTGGGCTTGCTGTGAGAGGATGACTT 1693

QY 601 GCACCTGGACAGCACTGATGTCACCTTGGAACCCCTTCAGACCAAGCTAACCTCCAGACA 660
DB 1694 GCACCTGGACAGCACTGATGTCACCTTGGAACCCCTTCAGACCAAGCTAACCTCCAGACA 1753
QY 661 GACAGATGGACCAAGACCAACGTCGAATATATCCAAATGTTAAATGTGAGTTTACCG 720
DB 1754 GACAGATGGACCAAGACCAACGTCGAATATATCCAAATGTTAAATGTGAGTTTACCG 1813
QY 721 CCTAGCTATGGACCTGCTGCTCTTCTGTCAGGAATCATGGGGGTATGACTGCTCCCA 780
DB 1814 CCTAGCTATGGACCTGCTGCTCTTCTGTCAGGAATCATGGGGGTATGACTGCTCCCA 1873
QY 781 ACCCTGTGGGCTGTATGACCAAGCTCAGGTAATCTCCCACTGGGGGCTGTGCCCTCCCT 840
DB 1874 ACCCTGTGGGCTGTATGACCAAGCTCAGGTAATCTCCCACTGGGGGCTGTGCCCTCCCT 1933
QY 841 GGGACGGTTCCTGGGACGCCCATCACTGTGTTCATATGATGTGAGATGTAGCTTAAGC 900
DB 1934 GGGACGGTTCCTGGGACGCCCATCACTGTGTTCATATGATGTGAGATGTAGCTTAAGC 1993
QY 901 CCTGCTGCTGCTGCTGCAATGCAATGCCACAGAGCGGTGGGGGCTGGTGGGGAATGCCA 960
DB 1994 CCTGCTGCTGCTGCTGCAATGCAATGCCACAGAGCGGTGGGGGCTGGTGGGGAATGCCA 2053
QY 961 TCGTGAAGTGTCTCTCACTTATGATGTGAGACAGAGACTTGGGGGAGATGCTCCAGGA 1020
DB 2054 TCGTGAAGTGTCTCTCACTTATGATGTGAGACAGAGACTTGGGGGAGATGCTCCAGGA 2113
QY 1021 TGTGGGTATTTCTGATCTGAGGAGGCTATCTTGAATCTCCGACAGGGGACACTCCCA 1080
DB 2114 TGTGGGTATTTCTGATCTGAGGAGGCTATCTTGAATCTCCGACAGGGGACACTCCCA 2173
QY 1081 GCCAGCCAGGGGTGAGGGGAGAGTGAACCTGAGATGAGCCAAAGCTGGGCTGAG 1140
DB 2174 GCCAGCCAGGGGTGAGGGGAGAGTGAACCTGAGATGAGCCAAAGCTGGGCTGAG 2233
QY 1141 GGAGCAGTGTGTTGAGCCAGAGACTTGGGGGAGGAGTGGGGGCTTTCTGCTGCT 1200
DB 2234 GGAGCAGTGTGTTGAGCCAGAGACTTGGGGGAGGAGTGGGGGCTTTCTGCTGCT 2293
QY 1201 CATTTGCTTCAATGAAGCTTCAAGAGCCAAACCAAGGCTTTCCCTCTCGAGT 1260
DB 2294 CATTTGCTTCAATGAAGCTTCAAGAGCCAAACCAAGGCTTTCCCTCTCGAGT 2353
QY 1261 TTGAATATCCAAATCTTTTGTACTTCTGTGTGTTAAATGTGTTATTTTGTAAAAAT 1320
DB 2354 TTGAATATCCAAATCTTTTGTACTTCTGTGTGTTAAATGTGTTATTTTGTAAAAAT 2413
QY 1321 AAAATATAATTTAGTTAATAAATGATGTTTTCACAGCAAACTCTTCCT 1368
DB 2414 AAAATATAATTTAGTTAATAAATGATGTTTTCACAGCAAACTCTTCCT 2461

RESULT 5

ID ABN94640/c

ABN94640 standard; DNA; 427 BP.

ABN94640;

13-AUG-2002 (first entry)

Gene #1138 used to diagnose liver cancer.

Gene; liver cancer; ds; hepatocellular carcinoma; hepatotropic;
metastatic liver tumor; cytostatic; expression profile; disease state;
disease progression; drug toxicity; drug efficacy; drug metabolism.

Homo sapiens.

WO200229103-A2.

11-APR-2002.


```

Db      1109 GGGCTCAGCTGAC-----CAGCTGGGCTGCTGCTGACGAGCTGACT 1151
Qy      600 TGACATGGGAGCAGCTGATGTCACCTGGGAACCCCTGGACAGCAAGCTAACCCAGAC 659
Db      1152 -----TGATCCACCTGGGAACCCCTGGACAGCAAAAGCTAGCTCCAGAG 1197
Qy      660 AGACAGATGTGACAGAGAC-AAACGTGCAATTAATGCAATGTTAAATGTGAGTTTAC 718
Db      1198 CAATGATGTGACAGAGCAAAAGCTGCAATATGCAAAATGTTAAATGTGAGTTTGC 1257
Qy      719 AGCTTACCTATGAGACTGCTGCTCCTAGTCCAGAAATCATGGGGGTATGATGCTCTC 778
Db      1258 AGCTT-----CAGTCCCACTGCTGTTGAAACCCCATTTGCTCTC 1296
Qy      779 CAACCTGTGGGCTGTAAAGCA-AGCTCAGGCTAGTGTCCCTCCACCTGGGGGCTGTGCCCTC 837
Db      1297 TGAGAGTGTAGGCTGTAGAGCAGGCTCAGGCTGCT-----TA 1335
Qy      838 CCTGGAGACGGTTCCTGGGAGCCCATCATCTGTGTTCAATAGTGTGAGAAATGACTAA 897
Db      1336 ACTGGAGCAGTCCCGTGGGAG--CCCATTTACTGCAATTG-ATGCTTTGAGAAATGAGCAG 1393
Qy      898 AGCCCTGCTGCTGCTGCTGCATGCAATGACAGAGCGGCTGGGGCTGCTGGGACAAAT 957
Db      1394 AAC-----ACTGCTGCTGCATTAAGCACCCTGGGCG--AGAGCTGCTGGGAGCAAC 1443
Qy      958 CCATCGTGAAGTTCCTTCAGCTTATGCTGTGAGCAGAGAACTGGCGGGGATGCTCCA 1017
Db      1444 CAGTCTCAGAGTCTCTCTAGCTCAGCTCCGCTCCAAATGAGAGCGCGGAGTG--CGG 1501
Qy      1018 GGAATGTGGGATTTCTGTAACCTGGGAGGCTATCTCTGACCTCCGACAGGGGACACTCC 1077
Db      1502 AGATGTGATGAGAACCAAGCACTGGGAAGAGGCTCTGGGCTCTCTCCCTAGAGGTTGCTC 1561
Qy      1078 CAGGCCAGCCAGGAGGCTCAGGGGAGAGGTGACACCTCAGATGAGCCAAAGCTGGGT 1137
Db      1562 TAGGCCAGCCCGAGGCGCGGAGAGAGTGTGCTGATCATGTAGCCAAAGACTAGAGT 1621
Qy      1138 CAGGAGCAGGTGTGTTGAGCCAGCACTGGGGGGGGGTGGGGCGGGGCTTTCTG 1197
Db      1622 GAGGAGCAGATTCATTGAGCCAGGA-CTGGGGGTGGGGGTGGGGGCTCTCTG 1680
Qy      1198 CCTCATTTGCTTCAATGAAGCCTCAAGACAGCC-----1232
Db      1681 CCTCATTTGCTTCAATGAAGCAGGAGCAGCCGAGCAGGCTCTCTCCACTCTG 1740
Qy      1233 -----AAACCAAGGCTTCCCTCTCTGAGTTGAATATC 1269
Db      1741 AGGCCAGGCTCTCCCTCTCTGAGGCGAGGCTCTCCCTCTGAGTTGGCTAC 1800
Qy      1270 CAGAACTTTTGTACTTCTTGTGTTAATTTGTTATTTTAAATAAATAAATAA 1329
Db      1801 CAGAACTTTTAACTTCTGTTCAATTAATTTTAAATAAATAAATAAATAAATAA 1859
Qy      1330 TTAGTTAATAAATGATTT 1350
Db      1860 TCAATTAATAAATGATTT 1880

```

RESULT 7

AA578838 standard; cDNA; 1837 BP.

AA578838;

13-FEB-2002 (first entry)

DNA encoding novel human diagnostic protein #14642.

Human; chromosome mapping; gene mapping; gene therapy; forensic; food supplement; medical imaging; diagnostic; genetic disorder; ss.

```

OS      Homo sapiens.
XX      PN
XX      MO200175067-A2.
XX      PD
XX      11-OCT-2001.
XX      PF
XX      30-MAR-2001; 2001WO-US08631.
XX      PR
XX      31-MAR-2000; 2000US-0540217.
XX      PR
XX      23-AUG-2000; 2000US-0649167.
XX      PA
XX      (HYSE-) HYSEQ INC.
XX      PI
XX      Drmanac RT, Liu C, Tang YT;
XX      WPI; 2001-639362/73.
XX      DR
XX      P-PSDB; ABG14651.
XX      PT
XX      New isolated polynucleotide and encoded polypeptides, useful in
XX      PT      diagnostics, forensics, gene mapping, identification of mutations
XX      PT      responsible for genetic disorders or other traits and to assess
XX      PT      biodiversity.
XX      PS
XX      Claim 1; SEQ ID No 14642; 103pp; English.
XX      PS
XX      The invention relates to isolated polynucleotide (I) and
XX      CC      polypeptide (II) sequences. (I) is useful as hybridisation probes,
XX      CC      polymerase chain reaction (PCR) primers, oligomers, and for chromosome
XX      CC      and gene mapping, and in recombinant production of (II). The
XX      CC      polynucleotides are also used in diagnostics as expressed sequence tags
XX      CC      for identifying expressed genes. (I) is useful in gene therapy techniques
XX      CC      to restore normal activity of (II) or to treat disease states involving
XX      CC      (II). (II) is useful for generating antibodies against it, detecting or
XX      CC      quantitating a polypeptide in tissue, as molecular weight markers and as
XX      CC      a food supplement. (II) and its binding partners are useful in medical
XX      CC      imaging of sites expressing (II). (I) and (II) are useful for treating
XX      CC      disorders involving aberrant protein expression or biological activity.
XX      CC      The polypeptide and polynucleotide sequences have applications in
XX      CC      diagnostics, forensics, gene mapping, identification of mutations in
XX      CC      responsible for genetic disorders or other traits to assess biodiversity
XX      CC      and to produce other types of data and products dependent on DNA and
XX      CC      amino acid sequences. AA564197-AA594564 represent novel human
XX      CC      diagnostic coding sequences of the invention.
XX      CC      Note: The sequence data for this patent did not appear in the printed
XX      CC      specification, but was obtained in electronic format directly from WIPO
XX      CC      at ftp.wipo.int/pub/published_pct_sequences.
XX      SQ
XX      Sequence 1837 BP; 424 A; 540 C; 471 G; 402 T; 0 other;
XX
XX      Query Match      13.8%; Score 192.8; DB 23; Length 1837;
XX      Best Local Similarity 97.5%; Pred. No. 4,3e-35;
XX      Matches 228; Conservative 0; Mismatches 2; Indels 4; Gaps 4;

```


QY 1339 AAAATGATGTTTCACAGCAAACTCTTCCT 1368
DB 30 AAAATGATGTTTCACAGCAAACTCTTCCT 1
RESULT 10
ABV05669/c
ID ABV05669 standard; cDNA; 429 BP.
AC ABV05669;
XX
XX 13-SEP-2002 (first entry)
DE Human prostate expression marker cDNA 5660.
XX
XX Human; prostate cancer; cytostatic; carcinogen; pharmacodynamic marker;
KM pharmacogenomic marker; gene; ss.
OS Homo sapiens.
XX
XX MO200160860-A2.
XX
XX 23-AUG-2001.
XX
XX 20-FEB-2001; 2001WO-US05171.
XX
XX 17-FEB-2000; 2000US-183319P.
XX 16-MAR-2000; 2000US-189862P.
XX 25-MAY-2000; 2000US-207454P.
XX 09-JUN-2000; 2000US-211314P.
XX 18-JUL-2000; 2000US-219007P.
XX 13-DEC-2000; 2000US-255281P.
XX
XX (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.
XX
XX Schlegel R, Endege WO, Monahan JE;
XX
XX WPI; 2001-662795/76.
XX
XX Novel isolated nucleic acid molecule associated with cancerous state of
PT prostate cells and correlating with presence of prostate cancer, useful
PT for detecting presence of prostate cancer, stage of prostate cancer
XX
XX Claim 1; Page 948; 11750bp; English.
XX
XX The invention relates to an isolated nucleic acid molecule (I) comprising
CC a nucleotide sequence given in Tables 1-9 (ABV0010-ABV62213) of the
CC specification or its complement. (I) is useful for:
CC (a) assessing whether a patient is afflicted with prostate cancer;
CC (b) monitoring the progression of prostate cancer in a patient;
CC (c) assessing the efficacy of a test compound to inhibit prostate
CC cancer in a patient;
CC (d) assessing the efficacy of a therapy for inhibiting prostate cancer
CC in a patient;
CC (e) selecting a composition for inhibiting prostate cancer in a patient;
CC (f) assessing the prostate cell carcinogenic potential of a compound;
CC (g) determining whether prostate cancer has metastasized in a patient;
CC (h) assessing the aggressiveness or indolence of prostate cancer in a
CC patient;
CC (I) is also useful as a pharmacodynamic or pharmacogenomic marker.
XX
XX Sequence 429 BP; 121 A; 68 C; 61 G; 106 T; 73 other;
SQ
Query Match 4.2%; Score 59; DB 23; Length 429;
Best Local Similarity 48.2%; Pred. No. 0.00043;
Matches 107; Conservative 0; Mismatches 115; Indels 0; Gaps 0;
QY 1169 GGGCGCGGGGTGGGCGCGGCTTCTGCTCATTTGCTTCAATGAAGCCTCAAGC 1228
DB 259 GNGGNGGNCNTNAGNNTNTNCTGGGNGGNTCTTNGGGGNTTTTNNCNCNCAANTN 200
QY 1229 AGCCAAAACCAAGGCTTCCCTTCGAGTTTGAATATCCAGAAATCTTTTACTTCT 1288

DB 199 AGNTTTTTTTTNGNGGCCCTCTGTGNGGTTTTCCTCNCNNAATTTTNTNGNTTT 140
QY 1289 TGTGTGTAATGTTTATTTTGTAAATAAATAATAGTTAATTAATGATG 1348
DB 139 TTTCCTTTAANTTTTTTTTTTNGNCCCAAAAAAAAAAAAAAAAAAAAAA 80
QY 1349 TTCACAGCAAACTCTTCCTTAAATAAAAAAAAAAAAAAAAAAAAAA 1390
DB 79 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 38
RESULT 11
AAS78836
ID AAS78836 standard; cDNA; 370 BP.
AC AAS78836;
XX
XX 13-FEB-2002 (first entry)
DE
XX
XX DNA encoding novel human diagnostic protein #14640.
XX
XX Human; chromosome mapping; gene mapping; gene therapy; forensic;
KM food supplement; medical imaging; diagnostic; genetic disorder; ss.
XX
XX Homo sapiens.
XX
XX MO200175067-A2.
XX
XX 11-OCT-2001.
XX
XX 30-MAR-2001; 2001WO-US08631.
XX
XX 31-MAR-2000; 2000US-0540217.
XX 23-AUG-2000; 2000US-0649167.
XX
XX (HYSE-) HYSEQ INC.
XX
XX Drmanac RT, Liu C, Tang YF;
XX
XX WPI; 2001-639362/73.
XX
XX P-PSDB; ABG14649.
XX
XX New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity
XX
XX Claim 1; SEQ ID No 14640; 103pp; English.
XX
XX The invention relates to isolated polynucleotide (I) and
CC polypeptide (II) sequences. (II) is useful as hybridisation probe,
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC and gene mapping, and in recombinant production of (II). The
CC polynucleotides are also used in diagnostics as expressed sequence tags
CC for identifying expressed genes. (I) is useful in gene therapy techniques
CC to restore normal activity of (II) or to treat disease states involving
CC (II). (II) is useful for generating antibodies against it, detecting or
CC quantitating a polypeptide in tissue, as molecular weight markers and as
CC a food supplement. (II) and its binding partners are useful in medical
CC imaging of sites expressing (II). (I) and (II) are useful for treating
CC disorders involving aberrant protein expression or biological activity.
CC The polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC responsible for genetic disorders or other traits to assess biodiversity
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. AAS64197-AAS94564 represent novel human
CC diagnostic coding sequences of the invention.
CC Note: The sequence data for this patent did not appear in the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pat_sequences.
SQ Sequence 370 BP; 67 A; 126 C; 101 G; 76 T; 0 other;

Query Match 3.9%; Score 54.6; DB 23; Length 370;
Best Local Similarity 93.4%; Pred. No. 0.0043; 4; Indels 0; Gaps 0;
Matches 57; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 87 CCATCCCCCAGGAGCTACCGCTGCTGACCATCTACCAACACGAGAGCTGCTCTT 146
DB 309 CACGCGACAGGAGACTACCGCTGCTGACCATCTACCAACACGAGAGCTGCTCTT 368
QY 147 C 147
DB 369 C 369

RESULT 12
ABV14838/C
ID ABV14838 standard; cDNA; 400 BP.
XX
AC ABV14838;
XX
DT 13-SEP-2002 (first entry)
XX
DE Human prostate expression marker cDNA 14829.
XX
DE Human prostate cancer; cytostatic; carcinogen; pharmacodynamic marker;
XX
KW pharmacogenomic marker; gene; ss.
XX
OS Homo sapiens.
XX
PN WO200160860-A2.
XX
PD 23-AUG-2001.
XX
PF 20-FEB-2001; 2001WO-US05171.
XX
PR 17-FEB-2000; 2000US-183319P.
XX
PR 16-MAR-2000; 2000US-189862P.
XX
PR 25-MAY-2000; 2000US-207454P.
XX
PR 09-JUN-2000; 2000US-211314P.
XX
PR 18-JUL-2000; 2000US-219007P.
XX
PR 13-DEC-2000; 2000US-255281P.
XX
PA (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.
XX
PI Schlegel R, Endege WO; Monahan JE;
XX
PI WPI; 2001-662795/76.
XX
DR
XX
PT Novel isolated nucleic acid molecule associated with cancerous state of
PT prostate cells and correlating with presence of prostate cancer, useful
PT for detecting presence of prostate cancer, stage of prostate cancer
XX
PS Claim 1; Page 2482; 11750pp; English.
XX
XX The invention relates to an isolated nucleic acid molecule (I) comprising
CC a nucleotide sequence given in Tables 1-9 (ABV00010-ABV62213) of the
CC specification or its complement. (I) is useful for:
CC (a) assessing whether a patient is afflicted with prostate cancer;
CC (b) monitoring the progression of prostate cancer in a patient;
CC (c) assessing the efficacy of a test compound to inhibit prostate
CC cancer in a patient;
CC (d) assessing the efficacy of a therapy for inhibiting prostate cancer
CC in a patient;
CC (e) selecting a composition for inhibiting prostate cancer in a patient;
CC (f) assessing the prostate cell carcinogenic potential of a compound;
CC (g) determining whether prostate cancer has metastasized in a patient;
CC (h) assessing the aggressiveness or indolence of prostate cancer in a
CC patient;
CC (I) is also useful as a pharmacodynamic or pharmacogenomic marker.
XX
SQ Sequence 400 BP; 101 A; 77 C; 82 G; 140 T; 0 other;

Query Match 3.7%; Score 51; DB 23; Length 400;
Best Local Similarity 56.1%; Pred. No. 0.031;

Matches 96; Conservative 0; Mismatches 75; Indels 0; Gaps 0;

QY 1220 CCTCAAGCAGCCAAACACAGCTTTCCCTCTCGGTTGATATCCAGATCTT 1279
DB 179 CCCCAATTAAGATTTTCTTTTGGGAGCCCCCTTGTTGTTTCCCTCCGAAATTTG 120
QY 1280 TGTACTCTTCTGTTGGTTAAATGTTTATTTTGTAAAAAATAAATTAGTTAATA 1339
DB 119 GTAAGTTTCTCTTTAATTTTCTTTTGGCCCAAAAAAATTTTAAAAA 60
QY 1340 AATGATGTTTACAGCAACACTTCCCTAAATTTTAAAAAATTTTAAAAA 1390
DB 59 AA 9

RESULT 13
ABV38926/C
ID ABV38926 standard; cDNA; 436 BP.
XX
AC ABV38926;
XX
DT 16-SEP-2002 (first entry)
XX
DE Human prostate expression marker cDNA 38917.
XX
DE Human prostate cancer; cytostatic; carcinogen; pharmacodynamic marker;
XX
KW pharmacogenomic marker; gene; ss.
XX
OS Homo sapiens.
XX
PN WO200160860-A2.
XX
PD 23-AUG-2001.
XX
PF 20-FEB-2001; 2001WO-US05171.
XX
PR 17-FEB-2000; 2000US-183319P.
XX
PR 16-MAR-2000; 2000US-189862P.
XX
PR 25-MAY-2000; 2000US-207454P.
XX
PR 09-JUN-2000; 2000US-211314P.
XX
PR 18-JUL-2000; 2000US-219007P.
XX
PR 13-DEC-2000; 2000US-255281P.
XX
PA (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.
XX
PI Schlegel R, Endege WO; Monahan JE;
XX
PI WPI; 2001-662795/76.
XX
DR
XX
PT Novel isolated nucleic acid molecule associated with cancerous state of
PT prostate cells and correlating with presence of prostate cancer, useful
PT for detecting presence of prostate cancer, stage of prostate cancer
XX
PS Claim 1; Page 7910-7911; 11750pp; English.
XX
XX The invention relates to an isolated nucleic acid molecule (I) comprising
CC a nucleotide sequence given in Tables 1-9 (ABV00010-ABV62213) of the
CC specification or its complement. (I) is useful for:
CC (a) assessing whether a patient is afflicted with prostate cancer;
CC (b) monitoring the progression of prostate cancer in a patient;
CC (c) assessing the efficacy of a test compound to inhibit prostate
CC cancer in a patient;
CC (d) assessing the efficacy of a therapy for inhibiting prostate cancer
CC in a patient;
CC (e) selecting a composition for inhibiting prostate cancer in a patient;
CC (f) assessing the prostate cell carcinogenic potential of a compound;
CC (g) determining whether prostate cancer has metastasized in a patient;
CC (h) assessing the aggressiveness or indolence of prostate cancer in a
CC patient;
CC (I) is also useful as a pharmacodynamic or pharmacogenomic marker.
XX
SQ Sequence 436 BP; 104 A; 89 C; 96 G; 147 T; 0 other;

Query Match 3.7%; Score 51; DB 23; Length 436;
 Best Local Similarity 56.1%; Pred. No. 0.032;
 Matches 96; Conservative 0; Mismatches 75; Indels 0; Gaps 0;

Db 1220 CCTCAAGCAGCCAAACAGGCTTCCCTCCCTGAGTTGAATTCAGAACTTT 1279
 217 CCCCAATAGATTTTCTTTTGGGAGCCCTGTTGTTTCCCTCCCAATTTTG 158

Qy 1280 TGTACTCTCTGTTGTTAAATGTTTATTTTGAATAAATAATTAATTAATA 1339
 157 GTAGGTTTCTTTCTTTAAATTTTCTTTGCCCCCAAAAAAAAAAAAAAAAA 98

Db 1340 AAATGATGTTTCACAGCAACTCTTCCCTAAATAAAAAAAAAAAAAA 1390
 97 AA 47

RESULT 14
 ABV44714/C
 ID ABV44714 standard; cDNA; 436 BP.
 AC ABV44714;
 XX
 DT 16-SEP-2002 (first entry)
 DE Human prostate expression marker cDNA 44705.
 XX
 KM Human; prostate cancer; cytostatic; carcinogen; pharmacodynamic marker;
 KM pharmacogenomic marker; gene; ss.
 OS Homo sapiens.
 XX
 PN WO200160860-A2.
 XX
 PD 23-AUG-2001.
 XX
 PF 20-FEB-2001; 2001WO-US05171.
 XX
 PR 17-FEB-2000; 2000US-183319P.
 PR 16-MAR-2000; 2000US-189862P.
 PR 25-MAY-2000; 2000US-207454P.
 PR 09-JUN-2000; 2000US-211314P.
 PR 18-JUL-2000; 2000US-219007P.
 PR 13-DEC-2000; 2000US-255281P.
 XX
 PA (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.
 XX
 PI Schlegel R, Endege WO, Monahan JE;
 XX
 DR WPI; 2001-662795/76.
 XX
 PT Novel isolated nucleic acid molecule associated with cancerous state of
 PT prostate cells and correlating with presence of prostate cancer, useful
 PT for detecting presence of prostate cancer, stage of prostate cancer -
 XX
 PS Claim 1; Page 8866; 11750P; English.
 XX
 CC The invention relates to an isolated nucleic acid molecule (I) comprising
 CC a nucleotide sequence given in Tables 1-9 (ABV00010-ABV62213) of the
 CC specification or its complement. (I) is useful for:
 CC (a) assessing whether a patient is afflicted with prostate cancer;
 CC (b) monitoring the progression of prostate cancer in a patient;
 CC (c) assessing the efficacy of a test compound to inhibit prostate
 CC cancer in a patient;
 CC (d) assessing the efficacy of a therapy for inhibiting prostate cancer
 CC in a patient;
 CC (e) selecting a composition for inhibiting prostate cancer in a patient;
 CC (f) assessing the prostate cell carcinogenic potential of a compound;
 CC (g) determining whether prostate cancer has metastasized in a patient;
 CC (h) assessing the aggressiveness or indolence of prostate cancer in a
 CC patient;
 CC (I) is also useful as a pharmacodynamic or pharmacogenomic marker.
 XX

Sequence 436 BP; 104 A; 89 C; 96 G; 147 T; 0 other;

Query Match 3.7%; Score 51; DB 23; Length 436;
 Best Local Similarity 56.1%; Pred. No. 0.032;
 Matches 96; Conservative 0; Mismatches 75; Indels 0; Gaps 0;

Db 1220 CCTCAAGCAGCCAAACAGGCTTCCCTCCCTGAGTTGAATTCAGAACTTT 1279
 217 CCCCAATAGATTTTCTTTTGGGAGCCCTGTTGTTTCCCTCCCAATTTTG 158

Qy 1280 TGTACTCTCTGTTGTTAAATGTTTATTTTGAATAAATAATTAATTAATA 1339
 157 GTAGGTTTCTTTCTTTAAATTTTCTTTGCCCCCAAAAAAAAAAAAAAAAA 98

Db 1340 AAATGATGTTTCACAGCAACTCTTCCCTAAATAAAAAAAAAAAAAA 1390
 97 AA 47

RESULT 15
 ABQ67093/C
 ID ABQ67093 standard; DNA; 83391 BP.
 AC ABQ67093;
 XX
 DT 28-AUG-2002 (first entry)
 DE Human angiogenesis associated polynucleotide SEQ ID NO 123.
 XX
 KM Human; angiogenesis; methylation; eye disease; glaucoma; tumour;
 KM inflammation; rheumatoid arthritis; diabetic retinopathy; antileucers;
 KM macular degeneration; inflammatory bowel disease; Crohn's disease;
 KM antirheumatic; antiarthritic; antidiabetic; antipsoriatic;
 KM arteriosclerotic; ds.
 XX
 OS Homo sapiens.
 XX
 PN WO200246454-A2.
 XX
 PD 13-JUN-2002.
 XX
 PF 06-DEC-2001; 2001WO-BP14320.
 XX
 PR 06-DEC-2000; 2000DE-1061338.
 XX
 PA (EPIC-) EPIGENOMICS AG.
 XX
 PI Schacht O;
 XX
 DR WPI; 2002-500450/53.
 XX
 PT New nucleic acid fragments from chemically treated
 PT angiogenesis-associated genes, useful for determining methylation
 PT status, e.g. in diagnosis or treatment of cancer -
 XX
 PS Claim 1; SEQ ID NO 123; 41pp + Sequence Listing; German.
 XX
 CC The invention relates to a nucleic acid (I) comprising a segment of 18
 CC bases of chemically pretreated DNA of angiogenesis-associated genes (II)
 CC having sequences (ABQ67093-ABQ67178) or their complements. (I), also
 CC related oligomers, are used to evaluate the methylation status and/or
 CC single-nucleotide polymorphisms, in angiogenesis-related genes, for
 CC diagnosis and treatment of eye diseases, proliferative retinopathy,
 CC neovascular glaucoma, solid tumours, inflammation, rheumatoid arthritis,
 CC diabetic retinopathy, macular degeneration caused by neovascularisation,
 CC psoriasis, arteriosclerosis, inflammatory bowel diseases, ulcers and
 CC Crohn's disease.
 CC Note: The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences.
 XX
 SQ Sequence 83391 BP; 25113 A; 665 C; 16761 G; 40835 T; 17 other;

Query Match 3.7%; Score 51; DB 24; Length 83391;
 Best Local Similarity 54.5%; Pred. No. 0.21;
 Matches 102; Conservative 0; Mismatches 85; Indels 0; Gaps 0;

QY	1204	TTGCTTCAATGAAGCCTCAAGCAGCCAAACGAGGCTTTCCCCCTTCCTGAGTTG	1263
DB	31892	TTTCTACATCTAAATCTAAACACCCCTAAACACCCCATCCCAATTTTAAACCTA	31833
QY	1264	AATATCCAGAACTTTTGATCTTCTGTGTTAAATGTTATTTTGTAAAAATAAA	1323
DB	31832	AAAAATATTATTTTATTAATCTTAACCTAAATAAAAAATTAATCTTTTACAAAT	31773
QY	1324	ATAAATTAAGTTAATAAATGATGTTTCACAGCAACTTCCTTAAAAAATAAAA	1383
DB	31772	ACCTATTCTCTATAAAAAAATAATCACATATTCTTAATCTAAAAAATAATAA	31713
QY	1384	AAAAAA 1390	
DB	31712	AAAAAA 31706	

Search completed: June 29, 2003, 04:14:17
 Tpb time : 251 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 29, 2003, 03:52:18 ; Search time 58 Seconds
(without alignments)
7386.680 Million cell updates/sec

Title: US-09-989-919-15

Perfect score: 1397
Sequence: 1 ggtgtctgcacccgtaccgga.....aaaaaaaaaaagcgctc 1397

Scoring table: IDENTITY NUC
Gapop 10.0 Gapext 1.0

Searched: 441362 seqs, 153338381 residues

Total number of hits satisfying chosen parameters: 882724

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents NA:
1: /cgn2_6/prodata/1/ina/5A_COMB.seq:*
2: /cgn2_6/prodata/1/ina/5B_COMB.seq:*
3: /cgn2_6/prodata/1/ina/5A_COMB.seq:*
4: /cgn2_6/prodata/1/ina/5B_COMB.seq:*
5: /cgn2_6/prodata/1/ina/PCTUS_COMB.seq:*
6: /cgn2_6/prodata/1/ina/backfilseq.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	43.2	3.1	2223	1 US-08-257-073-4	Sequence 4, Appl 1
2	43	3.1	944	3 US-08-906-769-82	Sequence 82, Appl 1
3	43	3.1	944	3 US-08-906-616-82	Sequence 82, Appl 1
4	43	3.1	944	3 US-08-817-795-82	Sequence 82, Appl 1
5	43	3.1	944	3 US-08-639-075A-82	Sequence 82, Appl 1
6	43	3.1	944	4 US-09-012-431-82	Sequence 82, Appl 1
7	43	3.1	944	4 US-09-012-692-82	Sequence 82, Appl 1
8	43	3.1	944	4 US-08-906-613-82	Sequence 82, Appl 1
9	43	3.1	944	5 PCT-US95-1442A-82	Sequence 82, Appl 1
10	43	3.1	945	4 US-09-032-215-26	Sequence 26, Appl 1
11	43	3.1	1582	3 US-08-545-196B-10	Sequence 28, Appl 1
12	43	3.1	1582	3 US-08-545-196B-10	Sequence 10, Appl 1
13	43	3.1	1582	3 US-08-545-196B-10	Sequence 12, Appl 1
14	42.6	3.0	7218	1 US-07-867-106-2	Sequence 2, Appl 1
15	42	3.0	5852	1 US-07-867-106-2	Sequence 1, Appl 1
16	41.8	3.0	4203	2 US-08-866-757-1	Sequence 1, Appl 1
17	41.8	3.0	4203	2 US-08-866-757-1	Sequence 1, Appl 1
18	41.4	3.0	721	4 US-09-288-143-38	Sequence 38, Appl 1
19	41.4	3.0	6152	4 US-08-973-462-1	Sequence 183, App 1
20	41.4	3.0	37950	4 US-09-338-907-183	Sequence 183, App 1
21	41.4	3.0	37950	4 US-09-338-907-183	Sequence 183, App 1
22	40.8	2.9	198	1 US-08-330-108-16	Sequence 16, Appl 1
23	40.8	2.9	198	5 PCT-US92-10087-16	Sequence 16, Appl 1
24	40.4	2.9	144	1 US-08-702-344-26	Sequence 26, Appl 1
25	40.4	2.9	1474	4 US-08-821-994-64	Sequence 64, Appl 1
26	40.4	2.9	3138	1 US-07-867-106-4	Sequence 4, Appl 1
27	40	2.9	1683	4 US-09-347-803-11	Sequence 11, Appl 1

28	40	2.9	2447	2 US-09-014-969-14	Sequence 14, Appl 1
29	39.8	2.8	1129	4 US-09-227-357-40	Sequence 40, Appl 1
30	39.6	2.8	1114	4 US-09-152-060-41	Sequence 41, Appl 1
31	39.6	2.8	2323	4 US-09-149-476-24	Sequence 24, Appl 1
32	39.4	2.8	1558	1 US-08-455-550-7	Sequence 7, Appl 1
33	39.4	2.8	1813	4 US-09-071-224-3	Sequence 3, Appl 1
34	39.2	2.8	1159	4 US-09-410-464-14	Sequence 14, Appl 1
35	39	2.8	1454	4 US-09-372-422A-19	Sequence 19, Appl 1
36	39	2.8	2202	4 US-09-465-558-59	Sequence 59, Appl 1
37	38.8	2.8	1096	4 US-09-461-697-26	Sequence 26, Appl 1
38	38.8	2.8	1147	1 US-08-665-716-1	Sequence 1, Appl 1
39	38.6	2.8	1882	4 US-09-370-253-1	Sequence 1, Appl 1
40	38.4	2.7	1181	4 US-09-149-476-310	Sequence 310, App 1
41	38.4	2.7	1212	4 US-09-149-476-186	Sequence 186, App 1
42	38.4	2.7	2334	1 US-08-062-632-4	Sequence 4, Appl 1
43	38.2	2.7	1134	3 US-09-248-335-29	Sequence 29, Appl 1
44	38.2	2.7	1492	4 US-08-745-995A-25	Sequence 25, Appl 1
45	38.2	2.7	1492	4 US-08-745-995A-27	Sequence 27, Appl 1

ALIGNMENTS

RESULT 1
US-08-257-073-4
; Sequence 4, Application US/08257073
; Patent No. 5766597
; GENERAL INFORMATION:
; APPLICANT: Paolietti, Enzo
; APPLICANT: de Taisne, Charles
; APPLICANT: Tine, John A.
; TITLE OF INVENTION: MALARIA RECOMBINANT POXVIRUS VACCINE
; NUMBER OF SEQUENCES: 143
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Curtis, Morris & Safford, P.C.
; STREET: 530 Fifth Avenue, 25th Floor
; CITY: New York
; STATE: New York
; COUNTRY: UNITED STATES OF AMERICA
; ZIP: 10036
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/257,073
; FILING DATE: 09-JUN-1994
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/075,783
; FILING DATE: 11-JUN-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/852,305
; FILING DATE: 18-MAR-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/672,183
; FILING DATE: 20-MAR-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: Frommer, William S.
; REGISTRATION NUMBER: 25,506
; REFERENCE/DOCKET NUMBER: 454310-2570
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 840-3333
; TELEFAX: (212) 840-0712
; TELEX: 425066 CURTMS
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2223 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; US-08-257-073-4

Query Match 3.1%; Score 43.2; DB 1; Length 2223;
Best Local Similarity 63.5%; Pred. No. 0.039;
Matches 66; Conservative 0; Mismatches 38; Indels 0; Gaps 0;

QY 1286 TCTGTGTTAAATTTTATTTTGTAAATAAATAAATAGTTAATAAATGA 1345
DB 1488 TGTATGATGAAATTTATTTCTTTGAAAAAGATTAATGAAAAATATATACCTTA 1547

QY 1346 TGTTCACAGCAAACTTCTTAAAAA 1389
DB 1548 TGTGTACACCAACTCAATCAAAAAAACAATAATGAAA 1591

RESULT 2

US-08-906-769-82
Sequence 82, Application US/08906769
Patent No. 6077687

GENERAL INFORMATION:

APPLICANT: Grieve, Robert B.
APPLICANT: Rushlow, Keith E.
APPLICANT: Wu Hunter, Shirley
APPLICANT: Frank, Glenn R.
APPLICANT: Stiegler, Gary
APPLICANT: Gaines, Patrick J.
APPLICANT: Silver, Gary
TITLE OF INVENTION: FLEA PROTEASE PROTEINS, NUCLEIC ACID
TITLE OF INVENTION: MOLECULES AND USES THEREOF
NUMBER OF SEQUENCES: 190
CORRESPONDENCE ADDRESS:
ADDRESSEE: Sheridan Ross & McIntosh
STREET: 1700 Lincoln Street, Suite 3500
CITY: Denver
STATE: Colorado
COUNTRY: USA
ZIP: 80203

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent in Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/906,769

FILING DATE:

CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/639,075
FILING DATE: 24-APR-1996

ATTORNEY/AGENT INFORMATION:

NAME: Connell, Gary J.
REGISTRATION NUMBER: 32,020
REFERENCE/DOCKET NUMBER: 2618-25-C2
TELECOMMUNICATION INFORMATION:
TELEPHONE: (303) 863-9700
TELEFAX: (303) 863-0223
INFORMATION FOR SEQ. ID NO: 82:
SEQUENCE CHARACTERISTICS:
LENGTH: 944 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: CDNA

FEATURE:
NAME/KEY: CDS
LOCATION: 3..768
US-08-906-769-82

Query Match 3.1%; Score 43; DB 3; Length 944;
Best Local Similarity 58.0%; Pred. No. 0.027;
Matches 76; Conservative 0; Mismatches 55; Indels 0; Gaps 0;

QY 1260 TTGAATATCCAGAACTTTGTACTCTCTGTGTTGTAATGTTTATTTTGTAAAAA 1319
DB 1380 AAAAAAAAAA 1390

DB 814 TTTAGTATATAAATCCCTTGGATTCATGCAAAATATTTGTTATTTATTTATTC 873

QY 1320 TAAAAATAAATAGTTAATAAATAGTTTACAGCAAACTCTCCCAAAAAAAAAA 1379
DB 874 TTATTCAAACGAATGTATTAAGTAAATTAACAAATAAATATTTAGTTGCCAAAAA 933

QY 1380 AAAAAAAAAA 1390
DB 934 AAAAAAAAAA 944

RESULT 3

US-08-906-616-82
Sequence 82, Application US/08906616
Patent No. 6121035

GENERAL INFORMATION:

APPLICANT: Grieve, Robert B.
APPLICANT: Rushlow, Keith E.
APPLICANT: Wu Hunter, Shirley
APPLICANT: Frank, Glenn R.
APPLICANT: Stiegler, Gary
APPLICANT: Gaines, Patrick J.
APPLICANT: Silver, Gary
TITLE OF INVENTION: FLEA AMINOPEPTIDASE PROTEINS AND USES THEREOF
NUMBER OF SEQUENCES: 190
CORRESPONDENCE ADDRESS:
ADDRESSEE: Sheridan Ross P.C.
STREET: 1700 Lincoln Street, Suite 3500
CITY: Denver
STATE: Colorado
COUNTRY: USA
ZIP: 80203

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent in Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/906,616
FILING DATE: 05-AUG-1997
CLASSIFICATION: 536
ATTORNEY/AGENT INFORMATION:
NAME: Connell, Gary J.
REGISTRATION NUMBER: 32,020
REFERENCE/DOCKET NUMBER: 2618-25-C2-3
TELECOMMUNICATION INFORMATION:
TELEPHONE: (303) 863-9700
TELEFAX: (303) 863-0223
INFORMATION FOR SEQ. ID NO: 82:
SEQUENCE CHARACTERISTICS:
LENGTH: 944 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: CDNA

FEATURE:
NAME/KEY: CDS
LOCATION: 3..768
US-08-906-616-82

Query Match 3.1%; Score 43; DB 3; Length 944;
Best Local Similarity 58.0%; Pred. No. 0.027;
Matches 76; Conservative 0; Mismatches 55; Indels 0; Gaps 0;

QY 1260 TTGAATATCCAGAACTTTGTACTCTCTGTGTTAATGTTTATTTTGTAAAAA 1319
DB 814 TTTAGTATATAAATCCCTTGGATTCATGCAAAATATTTGTTATTTATTTATTC 873

QY 1320 TAAAAATAAATAGTTAATAAATAGTTTACAGCAAACTCTCCCAAAAAAAAAA 1379
DB 874 TTATTCAAACGAATGTATTAAGTAAATTAACAAATAAATATTTAGTTGCCAAAAA 933

QY 1380 AAAAAAAAAA 1390

Db 934 AAAAAAAAAA 944

RESULT 4

US-08-817-795-82
Sequence 82, Application US/08817795

GENERAL INFORMATION:
APPLICANT: Grieve, Robert B.
APPLICANT: Rushlow, Keith E.
APPLICANT: Hunter, Shirley Wu
APPLICANT: Frank, Glenn W.
APPLICANT: Health, Andrew W.
APPLICANT: Yamaka, Miles Yamana
APPLICANT: Arfsten, Ann
APPLICANT: Dale, Beverly
APPLICANT: Stiegler, Gary
TITLE OF INVENTION: USE OF PROTEASE INHIBITORS AND
TITLE OF INVENTION: PROTEASE VACCINES TO PROTECT ANIMALS FROM FLEA
TITLE OF INVENTION: INFESTATION, AND FLEA PROTEASE PROTEINS, NUCLEIC ACID
TITLE OF INVENTION: MOLECULES, AND USES THEREOF
NUMBER OF SEQUENCES: 119
CORRESPONDENCE ADDRESS:
ADDRESSEE: Sheridan Ross & McIntosh
STREET: 1700 Lincoln Street, Suite 3500
CITY: Denver
STATE: Colorado
COUNTRY: USA
ZIP: 80203

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/817,795

FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/US95/14442

FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Gary J. Connell
REGISTRATION NUMBER: 32,020
REFERENCE/DOCKET NUMBER:
TELEPHONE: (303) 863-9700

TELECOMMUNICATION INFORMATION:
TELEPHONE: (303) 863-0223

INFORMATION FOR SEQ ID NO: 82:
SEQUENCE CHARACTERISTICS:
LENGTH: 944 base pairs

TYPE: nucleic acid
STRANDEDNESS: single

TOPOLOGY: linear
MOLECULE TYPE: CDNA

FEATURE:
NAME/KEY: CDS

LOCATION: 3..768

US-08-817-795-82

Query Match 3.1%; Score 43; DB 3; Length 944;
Best Local Similarity 58.0%; Pred. No. 0.027;
Matches 76; Conservative 0; Mismatches 55; Indels 0; Gaps 0;

QY 1260 TTGAATATCCAGATCTTTGTTACTCTTGTGTTAAATGTTTATTTGTAATAAAA 1319

DB 814 TTTTATTAATAAATTAATCTTTGATTCATGCAAAATATTTGTTATTTATTTATTTAC 873

QY 1320 TAAATATAATAGTAAATATAATGATGTTTCACAGCAACTCTTCCCTAAAAA 1379

DB 874 TTTATTCACACGATGTAATTAAGTAAATTAACATTAATAATGTTAGTGTGCCAAAAA 933

QY 1380 AAAAAAAAAA 1390
Db 934 AAAAAAAAAA 944

RESULT 5

US-08-639-075A-82
Sequence 82, Application US/08639075A

GENERAL INFORMATION:
APPLICANT: Grieve, Robert B.
APPLICANT: Rushlow, Keith E.
APPLICANT: Wu Hunter, Shirley
APPLICANT: Frank, Glenn R.
APPLICANT: Stiegler, Gary
APPLICANT: Gaines, Patrick J.
APPLICANT: Silver, Gary
TITLE OF INVENTION: FLEA PROTEASE PROTEINS, NUCLEIC ACID
TITLE OF INVENTION: MOLECULES AND USES THEREOF
NUMBER OF SEQUENCES: 190
CORRESPONDENCE ADDRESS:
ADDRESSEE: Sheridan Ross & McIntosh
STREET: 1700 Lincoln Street, Suite 3500
CITY: Denver
STATE: Colorado
COUNTRY: USA
ZIP: 80203

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/639,075A
FILING DATE: 24-APR-1996

CLASSIFICATION: 424
ATTORNEY/AGENT INFORMATION:
NAME: Connell, Gary J.
REGISTRATION NUMBER: 32,020
REFERENCE/DOCKET NUMBER: 2618-25-C2

TELECOMMUNICATION INFORMATION:
TELEPHONE: (303) 863-9700

INFORMATION FOR SEQ ID NO: 82:
SEQUENCE CHARACTERISTICS:
LENGTH: 944 base pairs
TYPE: nucleic acid
STRANDEDNESS: single

TOPOLOGY: linear
MOLECULE TYPE: CDNA

FEATURE:
NAME/KEY: CDS

LOCATION: 3..768

US-08-639-075A-82

Query Match 3.1%; Score 43; DB 3; Length 944;
Best Local Similarity 58.0%; Pred. No. 0.027;
Matches 76; Conservative 0; Mismatches 55; Indels 0; Gaps 0;

QY 1260 TTGAATATCCAGATCTTTGTTACTCTTGTGTTAAATGTTTATTTGTAATAAAA 1319

DB 814 TTTTATTAATAAATTAATCTTTGATTCATGCAAAATATTTGTTATTTATTTATTTAC 873

QY 1320 TAAATATAATAGTAAATATAATGATGTTTCACAGCAACTCTTCCCTAAAAA 1379

DB 874 TTTATTCACACGATGTAATTAAGTAAATTAACATTAATAATGTTAGTGTGCCAAAAA 933

QY 1380 AAAAAAAAAA 1390

DB 934 AAAAAAAAAA 944

RESULT 6

US-09-012-431-82
Sequence 82, Application US/09012431
Patent No. 6180383
GENERAL INFORMATION:
APPLICANT: Griewe, Robert B.
Rushlow, Keith E.
Wu Hunter, Shirley
Frank, Glenn R.
Stiegler, Gary
Gaines, Patrick J.
Silver, Gary
TITLE OF INVENTION: FLEA PROTEASE PROTEINS, NUCLEIC ACID
MOLECULES AND USES THEREOF
NUMBER OF SEQUENCES: 190
CORRESPONDENCE ADDRESS:
ADDRESSEE: Sheridan Rose & McIntosh
STREET: 1700 Lincoln Street, Suite 3500
CITY: Denver
STATE: Colorado
COUNTRY: USA
ZIP: 80203
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/012,431
FILING DATE: 23-Jan-1998
CLASSIFICATION: <Unknown>
PRIORITY APPLICATION DATA:
APPLICATION NUMBER: US 08/639,075
FILING DATE: <Unknown>
ATTORNEY/AGENT INFORMATION:
NAME: Connell, Gary J.
REGISTRATION NUMBER: 32,020
REFERENCE/DOCKET NUMBER: 2618-25-C2
TELECOMMUNICATION INFORMATION:
TELEPHONE: (303) 863-9700
TELEFAX: (303) 863-0223
INFORMATION FOR SEQ ID NO: 82:
SEQUENCE CHARACTERISTICS:
LENGTH: 944 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
FEATURE:
NAME/KEY: CDS
LOCATION: 3..768
SEQUENCE DESCRIPTION: SEQ ID NO: 82:
US-09-012-431-82
Query Match 3.1%; Score 43; DB 4; Length 944;
Best Local Similarity 58.0%; Pred. No. 0.027;
Matches 76; Conservative 0; Mismatches 55; Indels 0; Gaps 0;
QY 1260 TTGAATATCCAGATCTTTTGTACTCTTGTGGTTAAATGTTTATTTTGTAAAAA 1319
DB 814 TTTAGATATATAATATCTTTGATTCATGCAAAATATTTTGTATTATTTATTAC 873
QY 1320 TAAATAAATAATGTTAATAAATGATGTTTCACAGCAACTCTCCCTAAAAA 1379
DB 874 TTTATTCAAACGATGATTAAGATTAACAATAAATGTTAGTGGCCAAAAA 933
QY 1380 AAAAAAAAAA 1390
DB 934 AAAAAAAAAA 944

RESULT 7
US-09-012-692-82
Sequence 82, Application US/09012692

Patent No. 6214579
GENERAL INFORMATION:
APPLICANT: Griewe, Robert B.
Rushlow, Keith E.
Wu Hunter, Shirley
Frank, Glenn R.
Stiegler, Gary
Gaines, Patrick J.
Silver, Gary
TITLE OF INVENTION: FLEA PROTEASE PROTEINS, NUCLEIC ACID
MOLECULES AND USES THEREOF
NUMBER OF SEQUENCES: 190
CORRESPONDENCE ADDRESS:
ADDRESSEE: Sheridan Rose & McIntosh
STREET: 1700 Lincoln Street, Suite 3500
CITY: Denver
STATE: Colorado
COUNTRY: USA
ZIP: 80203
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/012,692
FILING DATE:
CLASSIFICATION:
PRIORITY APPLICATION DATA:
APPLICATION NUMBER: US 08/639,075
FILING DATE: 24-APR-1996
ATTORNEY/AGENT INFORMATION:
NAME: Connell, Gary J.
REGISTRATION NUMBER: 32,020
REFERENCE/DOCKET NUMBER: 2618-25-C2
TELECOMMUNICATION INFORMATION:
TELEPHONE: (303) 863-9700
TELEFAX: (303) 863-0223
INFORMATION FOR SEQ ID NO: 82:
SEQUENCE CHARACTERISTICS:
LENGTH: 944 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
FEATURE:
NAME/KEY: CDS
LOCATION: 3..768
US-09-012-692-82
Query Match 3.1%; Score 43; DB 4; Length 944;
Best Local Similarity 58.0%; Pred. No. 0.027;
Matches 76; Conservative 0; Mismatches 55; Indels 0; Gaps 0;
QY 1260 TTGAATATCCAGATCTTTTGTACTCTTGTGGTTAAATGTTTATTTTGTAAAAA 1319
DB 814 TTTAGATATATAATATCTTTGATTCATGCAAAATATTTTGTATTATTTATTAC 873
QY 1320 TAAATAAATAATGTTAATAAATGATGTTTCACAGCAACTCTCCCTAAAAA 1379
DB 874 TTTATTCAAACGATGATTAAGATTAACAATAAATGTTAGTGGCCAAAAA 933
QY 1380 AAAAAAAAAA 1390
DB 934 AAAAAAAAAA 944

RESULT 8
US-08-906-613-82
Sequence 82, Application US/08906613
Patent No. 6232096
GENERAL INFORMATION:
APPLICANT: Griewe, Robert B.

APPLICANT: Rushlow, Keith E.
APPLICANT: Wu Hunter, Shirley
APPLICANT: Frank, Glenn R.
APPLICANT: Stiegler, Gary
APPLICANT: Gaines, Patrick J.
APPLICANT: Silver, Gary
TITLE OF INVENTION: FLEA PROTEASE PROTEINS, NUCLEIC ACID
TITLE OF INVENTION: MOLECULES AND USES THEREOF
NUMBER OF SEQUENCES: 190
CORRESPONDENCE ADDRESS:
ADDRESSEE: Sheridan Ross & McIntosh
STREET: 1700 Lincoln Street, Suite 3500
CITY: Denver
STATE: Colorado
COUNTRY: USA
ZIP: 80203
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/906,613
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/639,075
FILING DATE: 24-APR-1996
ATTORNEY/AGENT INFORMATION:
NAME: Connell, Gary J.
REGISTRATION NUMBER: 32,020
REFERENCE/DOCKET NUMBER: 2618-25-C2
TELECOMMUNICATION INFORMATION:
TELEPHONE: (303) 863-9700
TELEFAX: (303) 863-0223
INFORMATION FOR SEQ ID NO: 82:
SEQUENCE CHARACTERISTICS:
LENGTH: 944 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
FEATURE:
NAME/KEY: CDS
LOCATION: 3..768
US-08-906-613-82

Query Match 3.1%; Score 43; DB 4; Length 944;
Best Local Similarity 58.0%; Pred. No. 0.027;
Matches 76; Conservative 0; Mismatches 55; Indels 0; Gaps 0;

QY 1260 TTGGATATCCAGAACTTTTGTGACTCTTGTTGTTAAATGTTTATTTTGTAAAAA 1319
DB 814 TTTAGTATATATAATCTTTGTGATTCAGCAAAATTTTGTATTTATTTATTTTAC 873
QY 1320 TAAATATTAATTAAGTATTAATGATGTTTCACAGCAACTCTTCCCTAAAAA 1379
DB 874 TTTATTCAAACGATGATTAAGTGAATTAACATATAAATGTTAGTGTGCCAAAAA 933
QY 1380 AAAAAAAAAA 1390
DB 934 AAAAAAAAAA 944

RESULT 9
PCT-US95-14442A-82
Sequence 82, Application PC/TUS9514442A
GENERAL INFORMATION:
APPLICANT: Grieve, Robert B.
APPLICANT: Rushlow, Keith E.
APPLICANT: Hunter, Shirley Wu
APPLICANT: Frank, Glenn R.
APPLICANT: Heath, Andrew W.

APPLICANT: Yamaka, Miles Yamanka
APPLICANT: Arfsten, Ann
APPLICANT: Dale, Beverly
APPLICANT: Stiegler, Gary
TITLE OF INVENTION: USE OF PROTEASE INHIBITORS AND
TITLE OF INVENTION: PROTEASE VACCINES TO PROTECT ANIMALS FROM FLEA
TITLE OF INVENTION: INFESTATION, AND FLEA PROTEASE PROTEINS, NUCLEIC ACID
NUMBER OF SEQUENCES: 119
CORRESPONDENCE ADDRESS:
ADDRESSEE: Sheridan Ross & McIntosh
STREET: 1700 Lincoln Street, Suite 3500
CITY: Denver
STATE: Colorado
COUNTRY: USA
ZIP: 80203
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US95/14442A
FILING DATE:
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Gary J. Connell
REGISTRATION NUMBER: 32,020
REFERENCE/DOCKET NUMBER:
TELECOMMUNICATION INFORMATION:
TELEPHONE: (303) 863-9700
TELEFAX: (303) 863-0223
INFORMATION FOR SEQ ID NO: 82:
SEQUENCE CHARACTERISTICS:
LENGTH: 944 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
FEATURE:
NAME/KEY: CDS
LOCATION: 3..768
PCT-US95-14442A-82

Query Match 3.1%; Score 43; DB 5; Length 944;
Best Local Similarity 58.0%; Pred. No. 0.027;
Matches 76; Conservative 0; Mismatches 55; Indels 0; Gaps 0;

QY 1260 TTGGATATCCAGAACTTTTGTGACTCTTGTTGTTAAATGTTTATTTTGTAAAAA 1319
DB 814 TTTAGTATATATAATCTTTGTGATTCAGCAAAATTTTGTATTTATTTATTTTAC 873
QY 1320 TAAATATTAATTAAGTATTAATGATGTTTCACAGCAACTCTTCCCTAAAAA 1379
DB 874 TTTATTCAAACGATGATTAAGTGAATTAACATATAAATGTTAGTGTGCCAAAAA 933
QY 1380 AAAAAAAAAA 1390
DB 934 AAAAAAAAAA 944

RESULT 10
US-09-032-215-26
Sequence 26, Application US/09032215
GENERAL INFORMATION:
PATENT NO. 6204010
APPLICANT: Stiegler, Gary L.
APPLICANT: Gaines, Patrick J.
TITLE OF INVENTION: FLEA PROTEASE PROTEINS, NUCLEIC
TITLE OF INVENTION: ACID MOLECULES, AND USES THEREOF
NUMBER OF SEQUENCES: 50
CORRESPONDENCE ADDRESS:
ADDRESSEE: Sheridan Ross P.C.

STREET: 1700 Lincoln Street, Suite 3500
CITY: Denver
STATE: Colorado
COUNTRY: U.S.A.
ZIP: 80203
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: ASCII DOS TEXT
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/032.215
FILING DATE: 27-FEB-1998
CLASSIFICATION: 536
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Connell, Gary J.
REGISTRATION NUMBER: 32,020
REFERENCE/DOCKET NUMBER: 2618-25-C6
TELECOMMUNICATION INFORMATION:
TELEPHONE: (303) 863-9700
TELEFAX: (303) 863-0223
INFORMATION FOR SEQ ID NO: 26:
SEQUENCE CHARACTERISTICS:
LENGTH: 945 nucleotides
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: CDNA
FEATURE:
NAME/KEY: CDS
LOCATION: 1..768
US-09-032-215-26

Query Match 3.1%; Score 43; DB 4; Length 945;
Best Local Similarity 58.0%; Pred. No. 0.027;
Matches 76; Conservative 0; Mismatches 55; Indels 0; Gaps 0;

QY 1260 TTGAATATCCAGAACTTTTGACTCTCTGTTGGTTAAATGTTTATTTTGTAAAAA 1319
DB 815 TTTTAGTAATAATAATCCCTTGATTCATGCAAAATTTTGTATTATTTATTTAC 874

QY 1320 TAAATATAATTAAGTATTAATAATGATGTTTCAAGCAACTCTCCCTAAAAAAA 1379
DB 875 TTTTATTCAGCAAGATGATTAAGTAACTTAACAATAAATGTTAGTGTGCCAAAAA 934

QY 1380 AAAAAAAAAA 1390
DB 935 AAAAAAAAAA 945

RESULT 11
US-09-032-215-28/C
Sequence 28, Application US/09032215
Patent No. 6204010
GENERAL INFORMATION:
APPLICANT: Stiegler, Gary L.
APPLICANT: Gaines, Patrick J.
TITLE OF INVENTION: FLEA PROTEASE PROTEINS, NUCLEIC
TITLE OF INVENTION: ACID MOLECULES, AND USES THEREOF
NUMBER OF SEQUENCES: 50
CORRESPONDENCE ADDRESS:
ADDRESSEE: Sheridan, Ross P.C.
STREET: 1700 Lincoln Street, Suite 3500
CITY: Denver
STATE: Colorado
COUNTRY: U.S.A.
ZIP: 80203
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: ASCII DOS TEXT
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/032.215
FILING DATE: 27-FEB-1998
CLASSIFICATION: 536
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Connell, Gary J.
REGISTRATION NUMBER: 32,020
REFERENCE/DOCKET NUMBER: 2618-25-C6
TELECOMMUNICATION INFORMATION:
TELEPHONE: (303) 863-9700
TELEFAX: (303) 863-0223
INFORMATION FOR SEQ ID NO: 28:
SEQUENCE CHARACTERISTICS:
LENGTH: 945 nucleotides
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: CDNA
US-09-032-215-28

Query Match 3.1%; Score 43; DB 4; Length 945;
Best Local Similarity 58.0%; Pred. No. 0.027;
Matches 76; Conservative 0; Mismatches 55; Indels 0; Gaps 0;

QY 1260 TTGAATATCCAGAACTTTTGACTCTCTGTTGGTTAAATGTTTATTTTGTAAAAA 1319
DB 131 TTTTAGTAATAATAATCCCTTGATTCATGCAAAATTTTGTATTATTTATTTAC 72

QY 1320 TAAATATAATTAAGTATTAATAATGATGTTTCAAGCAACTCTCCCTAAAAAAA 1379
DB 71 TTTTATTCAGCAAGATGATTAAGTAACTTAACAATAAATGTTAGTGTGCCAAAAA 12

QY 1380 AAAAAAAAAA 1390
DB 11 AAAAAAAAAA 1

RESULT 12
US-08-545-196B-10
Sequence 10, Application US/08545196B
Patent No. 6080577
GENERAL INFORMATION:
APPLICANT: MELKI, JUDITH
APPLICANT: MUNNICH, ARNOLD
TITLE OF INVENTION: SURVIVAL MOTOR NEURON (SMN) GENE: A GENE
TITLE OF INVENTION: FOR SPINAL MUSCULAR ATROPHY
NUMBER OF SEQUENCES: 65
CORRESPONDENCE ADDRESS:
ADDRESSEE: BIRCH, STEWART, KOLASCH AND BIRCH, LLP
STREET: PO BOX 747
CITY: FALLS CHURCH
STATE: VA
COUNTRY: USA
ZIP: 22040-0747
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/545.196B
FILING DATE: 19-OCT-1995
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: FARACT, C. J.
REGISTRATION NUMBER: 32,350
REFERENCE/DOCKET NUMBER: 2121-110P
TELECOMMUNICATION INFORMATION:

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic. - nucleic search, using sw model

Run on: June 29, 2003, 04:10:13 ; Search time 166 Seconds
(without alignments)
12492.616 Million cell updates/sec

Title: US-09-989-919-15

Perfect score: 1397
Sequence: 1 ggtgctgcacgtctaccgga.....aaaaaaaaaagcgctc 1397

Scoring table: IDENTITY NUC
Gap0 10.0, Gapext 1.0

Searched: 1055720 seqs, 742224136 residues

Total number of hits satisfying chosen parameters: 2111440

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database:

Published Applications NA:
1: /cgn2_6/prodata/1/pubpna/US07_PUBCOMB.seq:
2: /cgn2_6/prodata/1/pubpna/PCT_NEW_PUB.seq:
3: /cgn2_6/prodata/1/pubpna/US06_NEW_PUB.seq:
4: /cgn2_6/prodata/1/pubpna/US06_PUBCOMB.seq:
5: /cgn2_6/prodata/1/pubpna/US07_NEW_PUB.seq:
6: /cgn2_6/prodata/1/pubpna/PCTUS_PUBCOMB.seq:
7: /cgn2_6/prodata/1/pubpna/US08_NEW_PUB.seq:
8: /cgn2_6/prodata/1/pubpna/US08_PUBCOMB.seq:
9: /cgn2_6/prodata/1/pubpna/US09_NEW_PUB.seq:
10: /cgn2_6/prodata/1/pubpna/US09_PUBCOMB.seq:
11: /cgn2_6/prodata/1/pubpna/US10_NEW_PUB.seq:
12: /cgn2_6/prodata/1/pubpna/US10_PUBCOMB.seq:
13: /cgn2_6/prodata/1/pubpna/US60_NEW_PUB.seq:
14: /cgn2_6/prodata/1/pubpna/US60_PUBCOMB.seq:

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1397	100.0	1397	9 US-09-989-919-15	Sequence 15, Appl
2	464.2	33.2	470	9 US-09-989-919-14	Sequence 1138, Ap
3	420.8	30.1	427	10 US-09-980-107-1138	Sequence 1138, Ap
4	97	6.9	493	9 US-09-918-995-32213	Sequence 32213, A
5	55.2	4.0	428	10 US-09-960-352-10180	Sequence 10180, A
6	51	3.7	375	10 US-09-960-352-15014	Sequence 15014, A
7	50.6	3.6	435	10 US-09-834-975-451	Sequence 451, App
8	49.4	3.5	312	10 US-09-960-352-8414	Sequence 8414, Ap
9	49	3.5	361	9 US-10-198-846-8619	Sequence 8619, Ap
10	48.8	3.5	375	10 US-09-960-352-15014	Sequence 15014, A
11	48.8	3.5	424	10 US-09-960-352-11218	Sequence 11218, A
12	48.6	3.5	746	10 US-09-910-943-714	Sequence 714, App
13	48.4	3.5	298	10 US-09-960-352-1004	Sequence 1004, A
14	48.2	3.5	436	10 US-09-834-975-533	Sequence 533, App
15	48	3.4	424	9 US-10-198-846-2929	Sequence 2929, Ap
16	47.8	3.4	277	10 US-09-960-352-12673	Sequence 12673, A
17	47.8	3.4	312	10 US-09-960-352-8414	Sequence 8414, Ap
18	47.2	3.4	403	10 US-09-960-352-3319	Sequence 3319, Ap
19	47	3.4	638	9 US-10-198-846-8560	Sequence 8560, Ap

C 20	46.8	3.4	600	9 US-10-198-846-8434	Sequence 8434, Ap
C 21	46.8	3.4	1767	12 US-10-001-843-1	Sequence 1, Appl
C 22	46.6	3.3	416	10 US-09-960-352-4584	Sequence 4584, Ap
C 23	46.2	3.3	418	10 US-09-960-352-4845	Sequence 4845, Ap
C 24	46	3.3	242	10 US-09-960-352-3217	Sequence 3217, Ap
C 25	46	3.3	368	10 US-09-960-352-9843	Sequence 9843, Ap
C 26	46	3.3	448	9 US-10-198-846-2540	Sequence 2540, Ap
C 27	46	3.3	469	9 US-09-918-995-13017	Sequence 13017, A
C 28	46	3.3	2773	9 US-09-992-598-178	Sequence 178, App
C 29	46	3.3	2773	9 US-09-989-2928-178	Sequence 178, App
C 30	46	3.3	2773	9 US-10-063-547-33	Sequence 33, Appl
C 31	46	3.3	2773	9 US-09-989-734-178	Sequence 178, App
C 32	46	3.3	2773	9 US-09-990-444-178	Sequence 178, App
C 33	46	3.3	2773	9 US-09-989-730-178	Sequence 178, App
C 34	46	3.3	2773	9 US-09-990-436-178	Sequence 178, App
C 35	46	3.3	2773	9 US-09-991-181-178	Sequence 178, App
C 36	46	3.3	2773	9 US-09-993-687-178	Sequence 178, App
C 37	46	3.3	2773	9 US-09-989-734-178	Sequence 178, App
C 38	46	3.3	2773	9 US-09-997-653-178	Sequence 178, App
C 39	46	3.3	2773	9 US-10-174-590-149	Sequence 149, App
C 40	46	3.3	2773	9 US-10-176-758-149	Sequence 149, App
C 41	46	3.3	2773	9 US-10-063-616-33	Sequence 33, Appl
C 42	46	3.3	2773	9 US-10-175-737-149	Sequence 149, App
C 43	46	3.3	2773	9 US-09-993-667-178	Sequence 178, App
C 44	46	3.3	2773	9 US-10-063-502-33	Sequence 33, Appl
C 45	46	3.3	2773	9 US-10-173-706-149	Sequence 149, App

ALIGNMENTS

RESULT 1	US-09-989-919-15
Sequence 15, Application US/09989919	
Patent No. US2002016434A1	
GENERAL INFORMATION:	
APPLICANT: Macina, Roberto	
APPLICANT: Recipon, Hervé	
APPLICANT: Pluta, Jason	
APPLICANT: Ghosh, Malavika	
APPLICANT: Sun, Yongming	
APPLICANT: Liu, Chenghua	
TITLE OF INVENTION: Compositions and Methods Relating to Colon Specific Genes and Pr	
FILE REFERENCE: DEX-0289	
CURRENT APPLICATION NUMBER: US/09/989, 919	
CURRENT FILING DATE: 2001-11-21	
PRIOR APPLICATION NUMBER: 60/252,505	
PRIOR FILING DATE: 2000-11-22	
NUMBER OF SEQ ID NOS: 124	
SOFTWARE: PatentIn version 3.1	
SEQ ID NO 15	
LENGTH: 1397	
TYPE: DNA	
ORGANISM: Homo sapien	
US-09-989-919-15	
Query Match	100.0%; Score 1397; DB 9; Length 1397;
Best Local Similarity	100.0%; Pred. No. 0;
Matches 1397; Conservative	0; Mismatches 0; Indels 0; Gaps 0;
QY	1 GGTGTCACCTGTATCCAGAGCGGAGATGTGACAGACTCCACGAGCAAGCAGTAC 60
DB	1 GGTGTCACCTGTATCCAGAGCGGAGATGTGACAGACTCCACGAGCAAGCAGTAC 60
QY	61 CGAGTACAGTGTATCCAGAGCAAGCAATCCCGAAGACTACCGTGTGGCCATC 120
DB	61 CGAGTACAGTGTATCCAGAGCAAGCAATCCCGAAGACTACCGTGTGGCCATC 120
QY	121 CTAACACACGAGAGTGTCTCTTCACTGTTCACCTGCTGAGAGCTGTGATCTCG 180
DB	121 CTAACACACGAGAGTGTCTCTTCACTGTTCACCTGCTGAGAGCTGTGATCTCG 180
QY	181 TAGAGCCATGCCCAGTGTGGGCTTTGTGTGACCAACAGACACCTGACAGTGA 240

D	181	TGAGGCCATGCCAGTGTGGGCTTTGTGTGTCACCAACGACCTGTGACAGGTGA	240
O	241	GCAGTGGAGAAAGCCTTCCAAAGGAGATGGCAGACCTCTCTGGAGTTGATAGATAG	300
D	241	GCCAGTGGAGAAAGCCTTCCAAAGGAGATGGCAGACCTCTCTGGAGTTGATAGATAG	300
O	301	TGATCCCCCATCCGGAAGTCAGAGGGGGGTCTGAGGTGTATGAGAGAGAGCTTACGTCT	360
D	301	TGATCCCCCATTCGGAAGTCAGAGGGGGGTCTGAGGTGTATGAGAGAGAGCTTACGTCT	360
O	361	TCAAGCAGTCAAAATTAGAGGAATAGTCTTGGCTCCGAAAGAGAAACATCCAGCCCTG	420
D	361	TCAAGCAGTCAAAATTAGAGGAATAGTCTTGGCTCCGAAAGAGAAACATCCAGCCCTG	420
O	421	TTACCTCTCACTCTGCCCCCAGGTGGCAGCTGTCTTTTTCAGACTGAGATGAGCC	480
D	421	TTACCTCTCACTCTGCCCCCAGGTGGCAGCTGTCTTTTTCAGACTGAGATGAGCC	480
O	481	AAGTGTCCCTGATTCCCAAGAACCAATATGTGAAAGCCTCTGGCTGACCTATCTGAG	540
D	481	AAGTGTCCCTGATTCCCAAGAACCAATATGTGAAAGCCTCTGGCTGACCTATCTGAG	540
O	541	GGCTGGCTGACCACTGACTATCCTCAGCAGCTGGGCTTGCTGCTGAGAGGAGTACTT	600
D	541	GGCTGGCTGACCACTGACTATCCTCAGCAGCTGGGCTTGCTGCTGAGAGGAGTACTT	600
O	601	GCACTGGCAGCCTGACATGTCACTCTGGAAACCCCTGACAGACAAAGCTAACTCCACAGCA	660
D	601	GCACTGGCAGCCTGACATGTCACTCTGGAAACCCCTGACAGACAAAGCTAACTCCACAGCA	660
O	661	GACAGATGTGACCAAGACAAACGTGCATATATGCCAATGTTAAATGTGACTTTACCG	720
D	661	GACAGATGTGACCAAGACAAACGTGCATATATGCCAATGTTAAATGTGACTTTACCG	720
O	721	CCTACATAGGAGACGTCTGGCTCCTAGTCCAGGAATCATAGGGGGATATGACTGCTCTCA	780
D	721	CCTACATAGGAGACGTCTGGCTCCTAGTCCAGGAATCATAGGGGGATATGACTGCTCTCA	780
O	781	AACCTGTGGGCTGTAGAGCAAGCTCAGGCTAGTCTTCCCACTGGGGGCTGTGCCCTCCT	840
D	781	AACCTGTGGGCTGTAGAGCAAGCTCAGGCTAGTCTTCCCACTGGGGGCTGTGCCCTCCT	840
O	841	GGAGAGGTTCCGTGGGCAAGCCCATCACTGTGTTAATAGTGTGAAATGTAGCTAAAGC	900
D	841	GGAGAGGTTCCGTGGGCAAGCCCATCACTGTGTTAATAGTGTGAAATGTAGCTAAAGC	900
O	901	CCCTGCTGTGTCTGTGCACATGCCACAGAGGGGTGGGGCTGGGGGGAACAATCA	960
D	901	CCCTGCTGTGTCTGTGCACATGCCACAGAGGGGTGGGGCTGGGGGGAACAATCA	960
O	961	TCGTGAGATGTTCTCTCAGCTTAGGTCTGACAGAGACTTGGCGGGGATGTCTCAGGA	1020
D	961	TCGTGAGATGTTCTCTCAGCTTAGGTCTGACAGAGACTTGGCGGGGATGTCTCAGGA	1020
O	1021	TGTGGTATTTCTGTAACCTGGGGAGCTATCTTGAACCTCCGACAGGGGACACTCCAG	1080
D	1021	TGTGGTATTTCTGTAACCTGGGGAGCTATCTTGAACCTCCGACAGGGGACACTCCAG	1080
O	1081	GCCAGCCCAAGGGGTCAAGGGGCAAGAGTGCACACTCAGACATGAGCAAGACTGGGGTCA	1140
D	1081	GCCAGCCCAAGGGGTCAAGGGGCAAGAGTGCACACTCAGACATGAGCAAGACTGGGGTCA	1140
O	1141	GGAGAGAGTGTGTTTGAGACCAGACCTGGGGCGGGGGTGGGGCGGGGCTTTCTGCTC	1200
D	1141	GGAGAGAGTGTGTTTGAGACCAGACCTGGGGCGGGGGTGGGGCGGGGCTTTCTGCTC	1200
O	1201	CATTGTCTTCAATAAAGCCTCAAGAGACCCAAACAGGCTTCCCCCTTCTCGAGT	1260
D	1201	CATTGTCTTCAATAAAGCCTCAAGAGACCCAAACAGGCTTCCCCCTTCTCGAGT	1260
O	1261	TTGAATATCCAGATCTTTTGTACTTCTGTGGTTAAATGTTATTTTGTAAAAAT	1320
D	1261	TTGAATATCCAGATCTTTTGTACTTCTGTGGTTAAATGTTATTTTGTAAAAAT	1320

Accession	Sequence	Length
D6	1261 TTGAATCCGAGATCTTTGTAACCTCTCTGTGGTAAATGGTTATTTTGTGAAAAAAT	1350
QY	1321 AAAATAAATTAGTATTAATAAATGATGTTTCACAGCAAACTCTCCCTAAAAA	1360
DB	1321 AAAATAAATTAGTATTAATAAATGATGTTTCACAGCAAACTCTCCCTAAAAA	1360
QY	1381 AAAAAAAAAAGCGGCTC	1397
DB	1381 AAAAAAAAAAGCGGCTC	1397

RESULT 2
US-09-989-919-14
; Sequence 14, Application US/09989919

```

? Patent No. US20020164344A1
? GENERAL INFORMATION:
?
? APPLICANT: Macina, Roberto
? APPLICANT: Recipon, Hervé
? APPLICANT: Pluta, Jason
? APPLICANT: Ghosh, Malavika
? APPLICANT: Sun, Yongming
? APPLICANT: Liu, Chenghua
? TITLE OF INVENTION: Compositions and Methods Relating to Colon Specific Genes and Proteins
? FILE-REFERENCE: DEX-0289
?
? CURRENT APPLICATION NUMBER: US/09/989,919
?
? CURRENT FILING DATE: 2001-11-21
? PRIOR APPLICATION NUMBER: 60/252,505
? PRIOR FILING DATE: 2000-11-22
? NUMBER OF SEQ ID NOS: 124
? SOFTWARE: PatentIn version 3.1
?
? SEQ ID NO 14
?
? LENGTH: 470
?
? TYPE: DNA
? ORGANISM: Homo sapien
?
? US-09-989-919-14

```

Query Match	33.2%	Score 464.2	DB 9	Length 470
Best Local Similarity	99.4%	Pred. No. 3e-114		
Matches 466	Conservative 0	Mismatches 3	Indels 0	Gaps 0

QY	567	CAGCAGCTGGGCTTGGCTGTGGAGGAGTACTTGGACTGGACACTGCATGTCACTGG	626
Db	2	CAGGTGCTGGGCTTGGCTGTGGAGGAGTACTTGGACTGGACACTGCATGTCACTGG	61
QY	627	GGAACCCCTGCAGACAAGCTTAACATCCAGACAGACAGATGTGACCAGGACAACAGTGC	686
Db	62	GGAACCCCTGCAGACAAGCTTAACATCCAGACAGACAGATGTGACCAGGACAACAGTGC	121
QY	687	AATATATGCCAAATGTTAAATGTATGATTTACACGCTTAAGCTATATGGAGTCTGGCTCTTA	746
Db	122	AATATATGCCAAATGTTAAATGTATGATTTACACGCTTAAGCTATATGGAGTCTGGCTCTTA	181
QY	747	GTCGAGGATCATGGGGATGATGACTGCTCTCCACACCTGTGGGCTTAAGCAAGCTCAG	806
Db	182	GTCGAGGATCATGGGGATGATGACTGCTCTCCACACCTGTGGGCTTAAGCAAGCTCAG	241
QY	807	GCTAGTCTCCCACTGGGGGCTGTGCCCTTCTCGGAGCGGTTCCGTGGGCGGCCCATC	866
Db	242	GCTAGTCTCCCACTGGGGGCTGTGCCCTTCTCGGAGCGGTTCCGTGGGCGGCCCATC	301
QY	867	ACTAGTTCATATAGTGTGAGAAATGATAGCTTAAAGCCCTGCTGCTGCTGTSCACATGCA	926
Db	302	ACTAGTTCATATAGTGTGAGAAATGATAGCTTAAAGCCCTGCTGCTGCTGTSCACATGCA	361
QY	927	CAGCAGGCGGTGGGGGCTGCGGTGGGGACAATCCATCGTGAAGATTTCTTCAGCTTAAGT	986
Db	362	CAGCAGGCGGTGGGGGCTGCGGTGGGGACAATCCATCGTGAAGATTTCTTCAGCTTAAGT	421
QY	987	CTGCACACGAGACTTGGCGGGGGATGTCCACAGATATGGGGATTTCTGT	1035
Db	422	CTGCACACGAGACTTGGCGGGGGATGTCCACAGATATGGGGATTTCTGT	470

RESULT 3
US-09-880-107-1138/c
Sequence 1138, Application US/09880107
Patent No. US20020142981A1
GENERAL INFORMATION:
APPLICANT: Horne, Darci T.
APPLICANT: Vockley, Joseph G.
APPLICANT: Scherf, Uwe
APPLICANT: Gene Logic, Inc.
TITLE OF INVENTION: Gene Expression Profiles in Liver Cancer
FILE REFERENCE: 44921-5028-WO
CURRENT APPLICATION NUMBER: US/09/880.107
CURRENT FILING DATE: 2001-06-14
PRIOR APPLICATION NUMBER: US 60/211,379
PRIOR FILING DATE: 2000-06-14
PRIOR APPLICATION NUMBER: US 60/237,054
NUMBER OF SEQ ID NOS: 3950
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 1138
LENGTH: 427
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
OTHER INFORMATION: Genbank Accession No. US20020142981A1 AA451877
US-09-880-107-1138

Query Match 30.1%; Score 420.8; DB 10; Length 427;
Best Local Similarity 99.5%; Pred. No. 1.2e-102;
Matches 422; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 936 GTGGGGGCTGCGTGGGGAGACATCCATCGAGAGTGTCTCTCAGCTTAGGTGACAGG 995
DB 427 GTGGGGGCTGCGTGGGGAGACATCCATCGAGAGTGTCTCTCAGCTTAGGTGACAGG 368
QY 996 AGACTTGGGGGGGATGCTCCAGAGATGTGGGTGATTTCTGTAAGCTGGAGGCTATCTTG 1055
DB 367 AGACTTGGGGGGGATGCTCCAGAGATGTGGGTGATTTCTGTAAGCTGGAGGCTATCTTG 308
QY 1056 ACTTCCCGACAGGGGACATCCCGAGGCGCCAGGGGGTCAAGGGGAGAGGTGACACCT 1115
DB 307 ACTTCCCGACAGGGGACATCCCGAGGCGCCAGGGGGTCAAGGGGAGAGGTGACACCT 248
QY 1116 CAGCATGAGCAAGACTGGGGGTGAGGAGAGAGGTGATTTGAGCCAGAGGCTGGGGCGG 1175
DB 247 CAGCATGAGCAAGACTGGGGGTGAGGAGAGAGGTGATTTGAGCCAGAGGCTGGGGCGG 188
QY 1176 GGGTGGGGCCGGGGGCTTTCTGCTCATTTGCTTCAATGAAGCCTCAAGAGCCAAA 1235
DB 187 GGGTGGGGCCGGGGGCTTTCTGCTCATTTGCTTCAATGAAGCCTCAAGAGCCAAA 128
QY 1236 ACCAGGCTTCCCTCCCTCCAGGTGTAATATCCGAATCTTTTGTACTCTCTGTGCT 1295
DB 127 ACCAGGCTTCCCTCCCTCCAGGTGTAATATCCGAATCTTTTGTACTCTCTGTGCT 68
QY 1296 TAAATGTTATTTTGTAAAAAATAAATTAATAGTTAATAAATGATGTTTACAG 1355
DB 67 TAAATGTTATTTTGTAAAAAATAAATTAATAGTTAATAAATGATGTTTACAG 8
QY 1356 CAAA 1359
DB 7 CAAA 4

RESULT 4
US-09-918-995-32213
Sequence 32213, Application US/09918995
Patent No. US20030073623A1
GENERAL INFORMATION:
APPLICANT: Hyseq, Inc.
TITLE OF INVENTION: NOVEL NUCLEIC ACID SEQUENCES OBTAINED
FROM VARIOUS CDNA LIBRARIES
FILE REFERENCE: 20411-756

CURRENT APPLICATION NUMBER: US/09/918,995
CURRENT FILING DATE: 2001-07-30
PRIOR APPLICATION NUMBER: US/09/235,076
PRIOR FILING DATE: 1999-01-20
NUMBER OF SEQ ID NOS: 38054
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 32213
LENGTH: 493
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: misc.feature
LOCATION: (1)...(493)
OTHER INFORMATION: n = A,T,C or G
US-09-918-995-32213

Query Match 6.9%; Score 97; DB 9; Length 493;
Best Local Similarity 100.0%; Pred. No. 8.6e-16;
Matches 97; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GGTGCTGACCTGTACCGGAGGGGCGAGTATCTGACAGACTCCAGCGCAAGCAGAGTAC 60
DB 397 GGTGCTGACCTGTACCGGAGGGGCGAGTATCTGACAGACTCCAGCGCAAGCAGAGTAC 456
QY 61 CGAGTACAGTGTATCCAGACAGCAGCATCCCGCAG 97
DB 457 CGAGTACAGTGTATCCAGACAGCAGCATCCCGCAG 493

RESULT 5
US-09-960-352-10180/c
Sequence 10180, Application US/09960352
Patent No. US20020137139A1
GENERAL INFORMATION:
APPLICANT: Warren, Wesley C.
APPLICANT: Tao, Nengbing
APPLICANT: Byatt, John C.
APPLICANT: Mathialagan, Nagappan
TITLE OF INVENTION: NUCLEIC ACID AND OTHER MOLECULES ASSOCIATED WITH LACTATION AND
FILE REFERENCE: 16511.006/37-21(10298)C
CURRENT APPLICATION NUMBER: US/09/960,352
CURRENT FILING DATE: 2001-09-24
NUMBER OF SEQ ID NOS: 15112
SEQ ID NO 10180
LENGTH: 428
TYPE: DNA
ORGANISM: Bos taurus
OTHER INFORMATION: Clone ID: 44-LIB188-026-Q1-E1-C12
US-09-960-352-10180

Query Match 4.0%; Score 55.2; DB 10; Length 428;
Best Local Similarity 70.5%; Pred. No. 0.00013;
Matches 117; Conservative 0; Mismatches 43; Indels 6; Gaps 3;

QY 903 CTGCTGCTGCTGTGACATGCGACAGAGGGGTGGGGCTGGCGTGGGGAGCAATCATC 962
DB 427 CTGCTGCTGCTGTGACATGCGATGCGATGCG--TACGAGGGGCGCGAGGGGAGCAATCGTC 370
QY 963 GTGAGG--TGTCTCTCAGCTTAGGTCTGAGAGAGAGACTTGGCGGGGATGCTCCAGGA 1020
DB 369 ATGAGGACAGTCTCTTAGCCAGCTCCGAGCGGAGAGATTGAA--GGGATGCTCCAGAA 312
QY 1021 TGTGGGTATTTCTGTACTGCGGAGGCTATCTTGAATCCCGAC 1066
DB 311 TGTGGGTATTTCTGTACTGCGGAGGCTATCTTGAATCCCGAC 266

RESULT 6
US-09-960-352-15014
Sequence 15014, Application US/09960352
Patent No. US20020137139A1
GENERAL INFORMATION:

FEATURE:
NAME/KEY: misc feature
LOCATION: 15, 21, 25, 26, 29, 30, 33, 58, 59, 242, 243, 245, 247, 249,
LOCATION: 266, 268, 279, 281, 282, 283, 285, 287, 289, 290, 296,
LOCATION: 297, 311, 312, 313, 324, 325, 326, 329, 331, 343, 344, 351,
LOCATION: 352, 353, 361
OTHER INFORMATION: n = A,T,C or G
US-10-198-846-8619

Query Match 3.5%; Score 49; DB 9; Length 361;
Best Local Similarity 49.7%; Pred. No. 0.0052;
Matches 85; Conservative 0; Mismatches 86; Indels 0; Gaps 0;

QY 1220 CTTCAAGCAGCCAAACAGGCTTCCCTTCCTCGAGTTGAATATCCAGATCTT 1279
DB 335 CCCGNGGNN 276
QY 1280 TGTACTTCTTGTTGTTAAATCTTTATTTTGTAAATAATAATTAATTAATTA 1339
DB 275 TTTTNTNN 216
QY 1340 AATGATGTTTCACAGCAACTCTCCCTAAANNNNNNNNNNNNNNNNNNNNN 1390
DB 215 AA 165

RESULT 10

US-09-960-352-15014/C
Sequence 15014, Application US/09960352
Patent No. US20020137139A1

GENERAL INFORMATION:
APPLICANT: Warren, Wesley C.
APPLICANT: Tao, Nengbing
APPLICANT: Byatt, John C.
APPLICANT: Mathialagan, Nagappan
TITLE OF INVENTION: NUCLEIC ACID AND OTHER MOLECULES ASSOCIATED WITH LACTATION AND
FILE REFERENCE: 16511.006/37-21(10298)C
CURRENT APPLICATION NUMBER: US/09/960,352
CURRENT FILING DATE: 2001-09-24
NUMBER OF SEQ ID NOS: 15112
SEQ ID NO 15014
LENGTH: 375
TYPE: DNA
ORGANISM: Bos taurus
OTHER INFORMATION: Clone ID: 64-LIB3058-048-Q1-K1-H8
US-09-960-352-15014

Query Match 3.5%; Score 48.8; DB 10; Length 375;
Best Local Similarity 58.1%; Pred. No. 0.006;
Matches 86; Conservative 0; Mismatches 62; Indels 0; Gaps 0;

QY 1243 TTCCCTCTCTCTCGAGTTGAATATCCAGATCTTTGTAATCTTGTGTTAAATG 1302
DB 246 TTTTTCATCTCTTTTAAAAAATAATATATATTTTAAAAAATAATATATTTT 187
QY 1303 TTTATTTTGTAAATAATAATAATAATAATAATAATAATAATAATAATAATA 1362
DB 186 TTTAAATTTAAATAATAATAATAATAATAATAATAATAATAATAATAATAATA 127
QY 1363 TTCCCTAAANNN 1390
DB 126 AA 99

RESULT 11

US-09-960-352-11218/C
Sequence 11218, Application US/09960352
Patent No. US20020137139A1
GENERAL INFORMATION:
APPLICANT: Warren, Wesley C.
APPLICANT: Tao, Nengbing
APPLICANT: Byatt, John C.

APPLICANT: Mathialagan, Nagappan
TITLE OF INVENTION: NUCLEIC ACID AND OTHER MOLECULES ASSOCIATED WITH LACTATION AND
FILE REFERENCE: 16511.006/37-21(10298)C
CURRENT APPLICATION NUMBER: US/09/960,352
CURRENT FILING DATE: 2001-09-24
NUMBER OF SEQ ID NOS: 15112
SEQ ID NO 11218
LENGTH: 424
TYPE: DNA
ORGANISM: Bos taurus
OTHER INFORMATION: Clone ID: 48-LIB3058-026-Q1-K1-D12
US-09-960-352-11218

Query Match 3.5%; Score 48.8; DB 10; Length 424;
Best Local Similarity 59.3%; Pred. No. 0.0065;
Matches 83; Conservative 0; Mismatches 57; Indels 0; Gaps 0;

QY 1251 TTCTCGAGTTGAATATCCAGATCTTTGTAATCTTGTGTTAAATGTTAATTT 1310
DB 166 TTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTT 107
QY 1311 TGTAAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATA 1370
DB 106 TTAATTTAAANNN 47
QY 1371 AA 1390
DB 46 AA 27

RESULT 12

US-09-910-943-714
Sequence 714, Application US/09910943
Patent No. US20020081610A1

GENERAL INFORMATION:
APPLICANT: Hemmati-Briyanlou, Ali
APPLICANT: Altman, Curtis
TITLE OF INVENTION: Assays and Materials for Embryonic Gene Expression
FILE REFERENCE: 7529/16148US1
CURRENT APPLICATION NUMBER: US/09/910,943
CURRENT FILING DATE: 2001-07-23
NUMBER OF SEQ ID NOS: 742
SOFTWARE: PatentIn version 3.1
SEQ ID NO 714
LENGTH: 746
TYPE: DNA
ORGANISM: Xenopus laevis
FEATURE:
NAME/KEY: misc feature
LOCATION: (1) (746)
OTHER INFORMATION: n may be a or g or c or t/u
US-09-910-943-714

Query Match 3.5%; Score 48.6; DB 10; Length 746;
Best Local Similarity 56.2%; Pred. No. 0.011; 70; Indels 0; Gaps 0;
Matches 90; Conservative 0; Mismatches 70; Indels 0; Gaps 0;

QY 1238 CAGGCTTCCCTCTCTCGAGTTGAATATCCAGATCTTTGTAATCTTGTGTTA 1297
DB 125 CAGGATTTTCCCTCCCTCATTTGTTTCCAAAGCAATTTTGTCTTATCTTGATTT 184
QY 1298 AATGTTTATTTTGTAAATAATAATAATAATAATAATAATAATAATAATA 1357
DB 185 AAATTTTATTTTCAATCAANNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 244
QY 1358 AACTCTTCCCTAAANNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 1397
DB 245 AA 284

RESULT 13
US-09-960-352-1004/C

Db. 143 AGGCACITTTTTTATTCTTTTGGATTAAATAATTGGCTTTGAAAAAATTTTTTTTTT 84

Search completed: June 29, 2003, 05:25:37
Job time : 168 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 29, 2003, 02:17:52 ; Search time 1458 Seconds
(without alignments)
15517.898 Million cell updates/sec

Title: US-09-989-919-15

Perfect score: 1397
Sequence: 1 ggtgctgcacccctaccgga.....aaaaaaaaaaagcgctc 1397

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 16154066 seqs, 8097743376 residues

Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Database :

EST.*
1: em_estda:*
2: em_esthum:*
3: em_estin:*
4: em_estnu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_hic:*
9: gb_est1:*
10: gb_est2:*
11: gb_hic:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estcom:*
17: gb_gss:*
18: em_gss_hum:*
19: em_gss_inv:*
20: em_gss_pln:*
21: em_gss_vrt:*
22: em_gss_fun:*
23: em_gss_mam:*
24: em_gss_mus:*
25: em_gss_other:*
26: em_gss_pro:*
27: em_gss_rtd:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	703.2	50.3	1039	14	BM927766 AGENCOURT
2	643.6	46.1	1038	14	BM927767 AGENCOURT
3	626.4	44.8	773	14	BM671384 UI-E-CKI-
4	605.8	43.4	917	14	BO654344 AGENCOURT
5	602.8	43.1	645	13	BM65930 UI-E-DXI-
6	593	42.4	610	13	BM68538 UI-E-CKI-

C	7	588.4	42.1	678	10	AM409897	AM409897 fh03a06.y
C	8	574.6	40.1	776	12	BE827019	BE827019 602749053
C	9	562.6	41.3	770	10	BE250262	BE250262 600943273
C	10	562.4	40.3	565	10	AM516922	AM516922 xp86g09.x
C	11	554	39.7	935	14	BO647652	BO647652 AGENCOURT
C	12	545.4	39.0	548	14	BM692817	BM692817 UI-E-CKI-
C	13	541.4	38.8	543	13	BM142660	BM142660 1926C02.x
C	14	533.4	38.2	823	14	BO888182	BO888182 AGENCOURT
C	15	528	37.8	529	9	A1669816	A1669816 t311f08.x
C	16	520	37.2	520	14	BO013425	BO013425 UI-E-BC1D
C	17	519.4	37.2	522	10	AM592430	AM592430 hf42c03.x
C	18	500.4	35.8	504	10	BE677292	BE677292 7083b01.x
C	19	499.8	35.8	504	9	A1808803	A1808803 w197d10.x
C	20	497	35.6	518	9	A1313395	A1313395 qp80a02.x
C	21	495.2	35.4	888	14	BO652861	BO652861 AGENCOURT
C	22	494.8	35.4	498	10	AM512747	AM512747 xt77a02.x
C	23	492.4	35.2	521	14	BM680899	BM680899 UI-E-EJO-
C	24	491.4	35.2	493	10	BE205791	BE205791 bd49b01.x
C	25	490.8	35.1	494	9	AA552700	AA552700 nk15f10.s
C	26	488.8	35.0	492	9	A1453409	A1453409 t337h12.x
C	27	487.2	34.9	543	14	BM726293	BM726293 UI-E-EJO-
C	28	481.4	34.5	999	14	BO068521	BO068521 AGENCOURT
C	29	479.4	34.3	481	10	BO027861	BO027861 ws62b06.x
C	30	478.4	34.2	504	10	A1357244	A1357244 qx63b03.x
C	31	477.8	34.2	504	10	AM411057	AM411057 fh10b03.y
C	32	475.8	34.1	490	9	A1581877	A1581877 as02e05.x
C	33	475.6	34.0	665	12	BF528878	BF528878 602643341
C	34	475.4	34.0	489	10	AM001523	AM001523 wu33h08.x
C	35	473	33.9	485	14	BO012938	BO012938 UI-E-BC1D
C	36	465	33.3	465	14	BM693225	BM693225 UI-E-CKI-
C	37	461.8	33.1	507	9	AA037424	AA037424 2K36C04.s
C	38	458	32.8	800	12	BE826158	BE826158 602750515
C	39	457.4	32.7	459	10	AM263720	AM263720 xq47e02.x
C	40	457.2	32.7	563	9	AA135720	AA135720 z111b08.s
C	41	452.4	32.4	454	10	AM058133	AM058133 wv83f07.x
C	42	450.4	32.2	452	9	A1694266	A1694266 wu11g08.x
C	43	449.2	32.2	845	13	BM041039	BM041039 603644776
C	44	446.4	32.0	458	10	AM151856	AM151856 xf71g11.x
C	45	437.4	31.3	497	9	AA129876	AA129876 z113c09.s

ALIGNMENTS

RESULT 1
LOCUS BM927766 1039 bp mRNA linear EST 12-MAR-2002
DEFINITION AGENCOURT 6729770 NIH_MGC_100 Homo sapiens CDNA clone IMAGE:5797598
5', mRNA sequence.
ACCESSION BM927766
VERSION BM927766.1 GI:19378145
KEYWORDS EST.
SOURCE Homo sapiens
ORGANISM human.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 1039)
NIH-MGC <http://mgc.nci.nih.gov/>
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgabbs-remail.nih.gov
Tissue Procurement: CGAP (Stanford)
CDNA Library Preparation: Rubin Laboratory
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LNL at:
<http://image.jnl.gov>
Plate: LNCM2024 row: d column: 15
High quality sequence stop: 655.
Location/Qualifiers
1..1039

FEATURES
source

```

/organism="Homo sapiens"
/db xref="taxon:9606"
/clone_1ib="NIH_MGC_100"
/tissue_type="hepatocellular carcinoma, cell line"
/lab host="DH10B (phage-resistant)"
/notes="Organ: liver; Vector: pORF7; Site 1: XhoI; Site 2: EcoRI; cDNA made by oligo-dt priming. Directionally cloned into EcoRI/XhoI sites using the following 5' adaptor: GGCACGAG(G). Size-selected >500bp for average insert size 1.8kb. Library constructed by Ling Hong in the laboratory of Gerald M. Rubin (University of California, Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and Superscript II RT (Life Technologies). Note: this is a NIH_MGC Library."

BASE COUNT      223 a      276 c      300 g      237 t      3 others
ORIGIN

Query Match      50.3%; Score 703.2; DB 14; Length 1039;
Best Local Similarity 86.6%; Pred. No. 4.7e-109;
Matches 836; Conservative 0; Mismatches 111; Indels 18; Gaps 5;

420 GTTACTCTGACCTGCCCCCAGGTGGCAGCTGCTTTTCAAGACTGATGAGC 479
72 GTACACCAACGACCACTGAGAGGTGGAGCTGCTTTTCAAGACTGATGAGC 131
480 CAAGTGTGCTGATCCCAACAGACCAATATGTAAAGGCTTGGCTGATCTATGTA 539
132 CAAGTGTGCTGATCCCAACAGACCAATATGTAAAGGCTTGGCTGATCTATGTA 191
540 GGGGTGGGCTGACGAGTGAATCTCCAGAGTGGGCTTGGCTGAGAGGAGTACT 599
192 GGGGTGGGCTGACGAGTGAATCTCCAGAGTGGGCTTGGCTGAGAGGAGTACT 251
600 TGCATGCGACACTGTCATGTCACTTGGAAACCCCTGACAGACAAAGTTAATCCAC 659
252 TGCATGCGACACTGTCATGTCACTTGGAAACCCCTGACAGACAAAGTTAATCCAC 311
660 AGACATGTGACGACGACAAAGTGAATATGCAATATGTAATGATGATGATGATGAT 719
312 AGACATGTGACGACGACAAAGTGAATATGCAATATGTAATGATGATGATGATGAT 371
720 GCGTATGATGAGGACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 779
372 GCGTATGATGAGGACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 431
780 AACCTGTGGGCTGTAAAGCAAGCTCAGGCTAAGTCTCCCACTGGGGCTGTGCCCTCC 839
432 AACCTGTGGGCTGTAAAGCAAGCTCAGGCTAAGTCTCCCACTGGGGCTGTGCCCTCC 491
840 TGGGACGCTTCCGTGGGACGCCCATCATCTGTGTTCAATGTTGAATGTTAGTAAAG 899
492 TGGGACGCTTCCGTGGGACGCCCATCATCTGTGTTCAATGTTGAATGTTAGTAAAG 551
900 CCCCTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 959
552 CCCCTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 611
960 ATGTGAGTGTCTTCTCAAGCTTAGGTCTGAGACAGAGACTTGGCGGGGAGTGTCCAG 1019
612 ATGTGAGTGTCTTCTCAAGCTTAGGTCTGAGACAGAGACTTGGCGGGGAGTGTCCAG 671
1020 ATGTGAGTGTCTTCTCAAGCTTAGGTCTGAGACAGAGACTTGGCGGGGAGTGTCCAG 1079
672 ATGTGAGTGTCTTCTCAAGCTTAGGTCTGAGACAGAGACTTGGCGGGGAGTGTCCAG 731
1080 GGCAGCAGCC--AGGGGTGAGGGGAGAGGTGACACCTGACAGCATGAGACAGAGTGGAGT 1137
732 NGCCAGCCCAAGGGGTGAGGGGAGAGGTGACACCTGACAGCATGAGACAGAGTGGAGT 791
1138 -----CAGGAGAGAGGTGTGTTGAGCCAGAGACTGGGGCGGGGGTGGGGCGGG 1188
792 TCAAGGAACCAAGGTGTGTTGAGCCAGAGACTGGGGCGGGGGTGGGGCGGGCGGGG 851

```

```

1189 GCGTTCTG--CCTCATTTCTTTCA--TGAAGCCTCAAGCAGCCAAACAGGCTT 1244
852 CCTTTCGCGCTCATTTCTTTCAATGATAAAACCTTCAAGCAGCCAAACAGGCTT 911
1245 TCCCGCTTCTC--GAGTTGAATATTCAGAACTTTTGTACTTCTTGTGTTAATT 1301
912 TCCCGCTTCTC--GAGTTGAATATTCAGAACTTTTGTACTTCTTGTGTTAATT 971
1302 GTTATTTTGTAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 1361
972 AAAATGTTTATTTTGTAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 1031
1362 CTTC 1366
1032 TCCCC 1036

RESULT 2
BM927767      1038 bp      mRNA      linear      EST 12-MAR-2002
LOCUS      BM927767
DEFINITION      AGENCOURT 6729786 NIH_MGC_100 Homo sapiens cDNA clone IMAGE:5797599
5', mRNA sequence.
ACCESSION      BM927767
VERSION      BM927767.1 GI:19378146
KEYWORDS      EST.
SOURCE      human.
ORGANISM      Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE      1 (bases 1 to 1038)
AUTHORS      NIH-MGC http://img.ncl.nih.gov/.
TITLE      National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL      Unpublished (1999)
COMMENT      Contact: Robert Strausberg, Ph.D.
Email: cgabbs-remail.nih.gov
Tissue Procurement: CGAP (Stanford)
CDNA Library Preparation: Rubin Laboratory
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLCM2024 row: d column: 16
High quality sequence stop: 476.
Location/Qualifiers
1..1038
/organism="Homo sapiens"
/db xref="taxon:9606"
/clone_1ib="NIH_MGC_100"
/tissue_type="hepatocellular carcinoma, cell line"
/lab host="DH10B (phage-resistant)"
/notes="Organ: liver; Vector: pORF7; Site 1: XhoI; Site 2: EcoRI; cDNA made by oligo-dt priming. Directionally cloned into EcoRI/XhoI sites using the following 5' adaptor: GGCACGAG(G). Size-selected >500bp for average insert size 1.8kb. Library constructed by Ling Hong in the laboratory of Gerald M. Rubin (University of California, Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and Superscript II RT (Life Technologies). Note: this is a NIH_MGC Library."

BASE COUNT      191 a      320 c      284 g      218 t      25 others
ORIGIN

Query Match      46.1%; Score 643.6; DB 14; Length 1038;
Best Local Similarity 88.6%; Pred. No. 5e-99;
Matches 748; Conservative 0; Mismatches 89; Indels 7; Gaps 5;

420 GTTACTCTGACCTTGGCCCCCAGGTGGCAGCTGCTTTTCAAGACTGATGAGC 479
72 GTACACCAACGACCACTGAGAGGTGGAGCTGCTTTTCAAGACTGATGAGC 131

```


OY		480	CAATGTGGTCCCTGATTCCCAACAACACCAATATGTAAAGGCCTCTGGACTGAACCTTA	539
Db		132	CBAETGGTCCCTGATTCACACAGAACAATAATGTAAAGGCCTCTGGACTGAACCTTA	191
OY		540	GAGGCTCGGCTGACCAGCTGAACATACTCAGACAGCTGGGCTTGCTGTGAGGGAGTAGCT	599
Db		192	GGGCTCGGCTGACCAGCTGAACATACTCAGACAGCTGGGCTTGCTGTGAGGGAGTAGCT	251
OY		600	TGCACGTGGCACACATGCACTGATCACTCTGGMAAACCCTGACAGCAAAGCTAACATCCAGAC	659
Db		252	TGCACGTGGCACACATGCACTGATCACTCTGGMAAACCCTGACAGCAAAGCTAACATCCAGAC	311
OY		660	AGACAGATGTGACACAGAGCAAAAGTGTGAATTAATGCCAAATGTTAAATGTGAGTTTACCA	719
Db		312	AGACAGATGTGACACAGAGCAAAAGTGTGAATTAATGCCAAATGTTAAATGTGAGTTTACCA	371
OY		720	GCCTAGCATATGGAGACTCTGGCTCTTAGTCCAGGAATCATAG- GGGTATGATCGCTCTC	778
Db		372	GCCTAGCATATGGAGACTCTGGCTCTTAGTCCAGGAATCATAGAGGAGTATGATCGCTCTC	431
OY		779	CAACCCCTGTGGGCTGTGAAGCAAGCTCAAGGCTAGTCTTCCCAGT- GGAGGCTGTGCCCTC	837
Db		432	CAACCCCTGTGGGCTGTGAAGCAAGCTCAAGGCTATCTCCCAGTGGGAGGCTGTGCCCTC	491
OY		838	CCTGGAGACGGTTCGGTGGGAGAGCCCATCACTGTTCTAATPATGTGAGAAATGTACTTAA	897
Db		492	CCTGGAGACGGTTCGGTGGGAGAGCCCATCACTGTTCTAATPATGTGAGAAATGTACTTAA	551
OY		898	AGCCCTTGCTCTGCTGCTGTGACATATGCAACAGCAGCGCGTGGGGGCTGCTGGGAGCAAT	957
Db		552	AGCCCTTGCTCTGCTGCTGTGACATATGCAACAGCAGCGCGTGGGGGCTGCTGGGAGCAAT	611
OY		958	CCATGTGTGAGTGTCTCTCAGCTTAGGTGTGACAGAGACCTTGGCGGGAGTGTCCA	1017
Db		612	CCATGTGTGAGTGTCTCTCAGCTTAGGTGTGACAGAGACCTTGGCGGGAGTGTCCA	671
OY		1018	GGATGTGGGTATTTCTGTACTCTGGGAGGCTATCTGTACCTCCCCACAGGGGACACTCC	1077
Db		672	GGATGTGGGTATTTCTGTACTCTGGGAGGCTATCTGTACCTCCCCACAGGGGAACTCTCC	731
OY		1078	CAGGCCAAGCCAGAGGGGTCTAGGGGGAGAG-- GGTGACACCTCAGAGATGAGCAAGACT- G3	1134
Db		732	CAGGCCAAGCCAGAGGGGGTATAGGGGCAAAAAGGGGCACACCTCTCACATGAGGCCAAGACTGG6	791
OY		1135	GGTCAAGGAGACAGGTGTGG-- TTTGAGCCAGAGACTTGGGGCGGGGTTGGGGCCCGGGCCT	1192
Db		792	GGTCAAGGAGACAGGTGTGGTTTANCAAGAGAACTTGGGGCCGGGGGGTGGGGCGCNNNC	851
OY		1193	TTTGCTCTCATTTTGCTTTCAATGAAGCTCTAAGACGCCAABAAACAGGCTTTTCCCCTT	1252
Db		852	CTTTTCCCCCTCCTCNTTTGCTTTTAAAGGGCGGCCCCNNNMCNACCAGATCCCCGGCTTT	911
OY		1253	CCTC 1256	
Db		912	CCCC 915	

RESULT 3
BM671384/c

LOCUS DEFINITION UI-E-Ck1-efk-m-06-0-UI_82 UI-E-Ck1 Homo sapiens cdna clone

ACCESSION BM671384

VERSION BM671384.1 GI:18981282

KEYWORDS EST.

SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
REFERENCE AUTHORS Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
TITLE Normalization and subtraction: two approaches to facilitate gene discovery

JOURNAL	Genome Res. 6 (9), 791-806 (1996)
MEDLINE	97044477
COMMENT	Contact: Soares, MB Institute for Data Discovery and Modeling

Program for Rat Gene Discovery and Mapping
University of Iowa
451 Eckstein Medical Research Building Iowa City, IA 52242, USA
Tel: 319 335 8250
Fax: 319 335 9565
Email: msoares@blue.weeg.uiowa.edu
Tissue Procurement: Dr. Gregg Hageman
cDNA Library preparation: Dr. M. Bento Soares, University of Iowa
cDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa
DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
Clone Distribution: Researchers may obtain clones from Research
Genetics (www.resgen.com).
The following repetitive elements were found in this cDNA
sequence: 1-59, >(TAAA)n#simple_repeat (matched complement)
Seq primer: M13 Forward
POLYA=Yes, www.resgen.com

FEATURES	location/Qualifiers
source	1. .773

```

BASE COUNT
ORIGIN
181 a      221 c      180 g      189 t      2 others
TAG_SEQ=GTCC"
TAG_TISSUE=Foveal and Macular Retina
Eye Institute (NEI).
used to prime the synthesis of first-strand cDNA contains
a library tag sequence that is located between the Not I
site and the (dT)18 tail. The sequence tag for this
library is GTCC. This library was created for the program
Gene Discovery in the Visual System, supported by Nationala
Eye Institute (NEI).
TAG_LIB=UI-E-CK1
TAG_TISSUE=Foveal and Macular Retina
/tissue_type="Retina Foveal and Macular"
/clone_lib="UI-E-CK1"
/organism="Homo sapiens"
/db_xref="taxon:9606"
/notes="Organ: eye; Vector: pT73-Pac (Pharmacia) with a
modified polylinker; Site_1: Eor I; Site_2: Not I;
UI-E-CK1 is a normalized cDNA library containing the
following tissue(s): Retina Foveal and Macular. The
library was constructed according to Bernaldo, Lennon and
 Soares, Genome Research, 6:791-806, 1996. First strand
cDNA synthesis was primed with an oligo-dT primer
containing a Not I site. Double stranded cDNA was ligated
to an Eor I adaptor, digested with Not I, and cloned
directionally into pT73-Pac vector. The oligonucleotide
used to prime the synthesis of first-strand cDNA contains
a library tag sequence that is located between the Not I
site and the (dT)18 tail. The sequence tag for this
library is GTCC. This library was created for the program
Gene Discovery in the Visual System, supported by Nationala
Eye Institute (NEI).
TAG_LIB=UI-E-CK1
TAG_TISSUE=Foveal and Macular Retina

```

Query Match	44.8%;	Score 626.4;	DB 14;	Length 773;
Best Local Similarity	98.2%;	Pred. No. 4,6e-96;		
Matches 644;	Conservative	0;	Mismatches 11;	Indels 1; Gaps 1
Qy	720	GCCTAGCTATGGAGACTGCTGCTCCTTAAGTCAGAGATCATGGGGTATGACTGCTTCC	779	
Dh	656	GCTTACTGATGGAGACTGCTGCTGCTCCTTAAGTCAGAGATCATGGGGTATGACTGCTTCC	597	
Qy	780	-AACCCGTGGGCTGTAAAGCAAGCTCAGGCTAGTCTCCCACTGGGGGCTGTGCCCTCC	838	
Dh	596	NAACCCGTGGGCTGTAAAGCAAGCTCAGGCTAGTCTCCCACTGGGGGCTGTGCCCTCC	537	
Qy	839	CTGGGACGGTTCCGTGGGAGCGCCCATCACTGTGTTCAATAGTGTGAGAAATGTAAGCTTAA	898	
Dh	536	CTGGGACGGTTCCGTGGGAGCGCCCATCACTGTGTTCAATAGTGTGAGAAATGTAAGCTTAA	477	
Qy	899	GCCCCCTGCTGCTGCTGCTGACATCCACAGAGGCGGTGGGGGCTGCTGGGGGACAATC	958	
Dh	476	GCCCCCTGCTGCTGCTGCTGACATCCACAGAGGCGGTGGGGGCTGCTGGGGGACAATC	417	
Qy	959	CATCGTGAGAGTTCCTCACAAGTTAGGTGTGGAGACAGAAATTGGCGGGGGAGTGTCCAG	101	
Dh	416	CATCGTGAGAGTTCCTCACAAGTTAGGTGTGGAGACAGAAATTGGCGGGGGAGTGTCCAG	357	

University of Iowa
451 Eckstein Medical Research Building Iowa City, IA 52242, USA
Tel: 319 335 8250
Fax: 319 335 9565
Email: mscares@blue.weeg.uiowa.edu

Tissue Procurement: Dr. Gregg Hageman
CDNA Library Preparation: Dr. M. Bento Soares, University of Iowa
CDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa
DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
Clone Distribution: Researchers may obtain clones from Research Genetics (www.resgen.com)
The following repetitive elements were found in this CDNA sequence: 1-59, >(TAAA)n#simple_repeat (matched compliment)
Seq primer: M13 Forward
POLYA=Yes.

FEATURES

source
1. 645
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="UI-E-DX1-agn-1-21-0-UI"
/clone_1ib="UI-E-DX1"
/tissue_type="fetal eyes"
/dev_stage="fetal"
/lab_host="DH10B (Life Technologies) (T1 phage resistant)"
/note="Organ: eye; Vector: pT73-Pac (Pharmacia) with a modified polylinker; Site 1: EcoR I; Site 2: Not I; UI-E-DX1 is a normalized CDNA library containing the following tissue(s): fetal eyes. The library was constructed according to Bonaldo, Lennon and Soares, Genome Research, 6:791-806, 1996. First strand cDNA synthesis was primed with an oligo-dT primer containing a Not I site. Double stranded cDNA was ligated to an EcoR I adaptor, digested with Not I, and cloned directionally into pT73-Pac vector. The oligonucleotide used to prime the synthesis of first-strand cDNA contains a library tag sequence that is located between the Not I site and the (dT)18 tail. The sequence tag for this library is AGAATCAAGA. This library was created for the program, Gene Discovery in the Visual System, supported by National Eye Institute (NEI).
TAG_LIB=UI-E-DX1
TAG_TISSUE=human fetal eyes
TAG_SEQ=AGAATCAAGA"
BASE COUNT 152 a 188 c 155 g 150 t
ORIGIN
Query Match 43.1%; Score 602.8; DB 13; Length 645;
Best Local Similarity 97.8%; Pred. No. 4,7e-92;
Matches 632; Conservative 0; Mismatches 12; Indels 2; Gaps 2;
730 GGGACTGCTGGCTCCAGTATCATGAGGAGTATGACTGCTCTCCAA-CCTGTG 788
645 GGGACTGCTGGCTCCAGTATCATGAGGAGTATGACTGCTCTCCAA-CCTGTG 586
789 GGGCTGAAGCAAGCTAGCTCTCCCACTGGGGGCTGTCCTCCCTGGGACG 848
585 GGGCTGAAGCAAGCTAGCTCTCCCACTGGGGGCTGTCCTCCCTGGGACG 526
849 TCCGTGGGCGACCCCATCACTGTGTTCAATAGTGTGAAGTGAAGCCCTGCTG 908
525 TCCGTGGGCGACCCCATCACTGTGTTCAATAGTGTGAAGTGAAGCCCTGCTG 466
909 CTGCTCTGACATGACAGAGGCGGTGGGGGCTGGGGGAGCAATTCATCTGGAG 968
465 CTGCTCTGACATGACAGAGGCGGTGGGGGCTGGGGGAGCAATTCATCTGGAG 407
969 TGTCTCTGACATGAGCTGACAGAGAGCTGGCGGGGAGTCTCCAGAGTGGGTG 1028
406 TGTCTCTGACATGAGCTGACAGAGAGCTGGCGGGGAGTCTCCAGAGTGGGTG 347
1029 ATTCTGTACTGGGAGGCTATCTTGACCTCCCGACAGGGGACACTCCAGGCCAGCC 1088
346 ATTCTGTACTGGGAGGCTATCTTGACCTCCCGACAGGGGACACTCCAGGCCAGCC 287

QY 1089 AGGGGTCAAGGGGAGAGTGACACCTCAGCATGACCAAGACTGGGGTTCAGGACAG 1148
DB 286 AGGGGTCAAGGGGAGAGTGACACCTCAGCATGACCAAGACTGGGGTTCAGGACAG 227
QY 1149 TGTGTTTGAAGCCAGGACCTGGGGGCGGGGCTGCTTCTGCTCATTTGCT 1208
DB 226 TGTGTTTGAAGCCAGGACCTGGGGGCGGGGCTGCTTCTGCTCATTTGCT 167
QY 1209 TTCAATGAAGCCTCAAGCAGCAACAGGCTTCCCTCTCCAGTTGAAT 1268
DB 166 TTCAATGAAGCCTCAAGCAGCAACAGGCTTCCCTCTCCAGTTGAAT 107
QY 1269 CCAGATCTTTGTACTCTTGTGTTGTTAAATGTTTAAATTAATTAATTA 1328
DB 106 CCAGATCTTTGTACTCTTGTGTTGTTAAATGTTTAAATTAATTAATTA 47
QY 1329 ATTAGTTAATTAATGATGTTTCAACAGCAAACTCTTCCCTAATAA 1374
DB 46 ATTAGTTAATTAATGATGTTTCAACAGCAAACTCTTCCCTAATAA 1

RESULT 6
LOCUS BM668538/c 610 bp mRNA linear EST 27-FEB-2002
DEFINITION UI-E-CK1-afm-c-23-0-UI.52 UI-E-CK1 Homo sapiens CDNA clone
ACCESSION UI-E-CK1-afm-c-23-0-UI 3', mRNA sequence.
VERSION BM668538.1 GI:18976369
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryote; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE 1 (bases 1 to 610)
AUTHORS Bonaldo,M.F., Lennon,G. and Soares,M.B.
TITLE Normalization and subtraction: two approaches to facilitate gene discovery
JOURNAL Genome Res. 6 (9), 791-806 (1996)
MEDLINE 97044477
COMMENT Contact: Soares, MB
Program for Rat Gene Discovery and Mapping
University of Iowa
451 Eckstein Medical Research Building Iowa City, IA 52242, USA
Tel: 319 335 8250
Fax: 319 335 9565
Email: mscares@blue.weeg.uiowa.edu

Tissue Procurement: Dr. Gregg Hageman
CDNA Library Preparation: Dr. M. Bento Soares, University of Iowa
CDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa
DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
Clone Distribution: Researchers may obtain clones from Research Genetics (www.resgen.com).
The following repetitive elements were found in this CDNA sequence: 32-80, >AT rich#low_complexity
Seq primer: M13 Forward
POLYA=Yes.

FEATURES

source
1. 610
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="UI-E-CK1-afm-c-23-0-UI"
/clone_1ib="UI-E-CK1"
/tissue_type="Retina Foveal and Macular"
/dev_stage="adult"
/lab_host="DH10B (Life Technologies) (T1 phage resistant)"
/note="Organ: eye; Vector: pT73-Pac (Pharmacia) with a modified polylinker; Site 1: EcoR I; Site 2: Not I; UI-E-CK1 is a normalized CDNA library containing the following tissue(s): Retina Foveal and Macular. The library was constructed according to Bonaldo, Lennon and Soares, Genome Research, 6:791-806, 1996. First strand cDNA synthesis was primed with an oligo-dT primer

containing a Not I site. Double stranded cDNA was ligated to an EcoR I adaptor, digested with Not I, and cloned directionally into pMT3-Pac Vector. The oligonucleotide used to prime the synthesis of first-strand cDNA contains a library tag sequence that is located between the Not I site and the (dr)18 tail. The sequence tag for this library is GTCC. This library was created for the program, Gene Discovery in the Visual System, supported by National Eye Institute (NEI).

TAG: LIB=UI-E-CXI
TAG_TISSUE=Foveal and Macular Retina
TAG_SEQ=GTCC

BASE COUNT 145 a 147 g 143 t 1 others
ORIGIN

Query Match 42.4%; Score 593; DB 13; Length 610;
Best Local Similarity 98.2%; Pred. No. 2.2e-90;
Matches 599; Conservative 0; Mismatches 11; Indels 0; Gaps 0;

765 TATGACTGCTCTCCACCTGCTGGGCTGTAGACAGCTAGCTCCCTCCACTGGG 824
610 TATGACTGCTCTCCACCTGCTGGGCTGTAGACAGCTAGCTCCCTCCACTGGG 551
825 GGCTGTGCCCCCTCCCTGAGCGGTTCCGTGGGAGCCCCCATCTGTGTTCAATAGTGTG 884
550 GGCTGTGCCCCCTCCCTGAGCGGTTCCGTGGGAGCCCCCATCTGTGTTCAATAGTGTG 491
885 AGAATGATGCTAAAGCCCTCTGCTGCTGTCACATCCACAGCAGGGGTGGGGGCT 944
490 AGAATGATGCTAAAGCCCTCTGCTGCTGTCACATCCACAGCAGGGGTGGGGGCT 431
945 GCGTGGGAGCAATCCATCTGAGTGTCTCTCAGCTAGGTGAGCAGGAGACTGGC 1004
430 GCGTGGGAGCAATCCATCTGAGTGTCTCTCAGCTAGGTGAGCAGGAGACTGGC 371
1005 GGGGATGCTCCAGATGTGGGTATCTGTACCTGGGAGGCTATCTGACCTCCCGA 1064
370 GGGGATGCTCCAGATGTGGGTATCTGTACCTGGGAGGCTATCTGACCTCCCGA 311
1065 CAGGGAGACATCCCGAGGCGAGCCGAGGGGTGAGGGGAGAGGTGACACTCAGCATGAG 1124
310 CAGGGAGACATCCCGAGGCGAGCCGAGGGGTGAGGGGAGAGGTGACACTCAGCATGAG 251
1125 CCAAGACTGGGGGTGAGGAGAGGTGTGTTGAGCAGGACCTGGGGCGGGGTGGGG 1184
250 CCAAGACTGGGGGTGAGGAGAGGTGTGTTGAGCAGGACCTGGGGCGGGGTGGGG 191
1185 CCGGGCTTTCTGCTCATTTGCTTCAATGAAGCTCAAGAGCCCAAGCAGGCTT 1244
190 CCGGGCTTTCTGCTCATTTGCTTCAATGAAGCTCAAGAGCCCAAGCAGGCTT 131
1245 TCCCTCTCTCGAGTTGATATCCAGATCTTTTGTACTCTGTGTGTTAAATGTT 1304
130 TCCCTCTCTCGAGTTGATATCCAGATCTTTTGTACTCTGTGTGTTAAATGTT 71
1305 TATTTTGTAAATAAATAAATAAATTAATTAATTAATTAATTAATTAATTAATTAAT 1364
70 TATTTTGTAAATAAATAAATAAATTAATTAATTAATTAATTAATTAATTAATTAAT 11
1365 CCTTAAAAA 1374
10 AAAAAAAAAA 1

RESULT 7
LOCUS AM409897/c
DEFINITION fh03d06.y1 NIH_MGC_17 Homo sapiens cDNA clone IMAGE:2960938 3',
678 bp mRNA linear EST 29-JUN-2000
AM409897
mRNA sequence.
ACCESSION AM409897
VERSION AM409897
KEYWORDS EST.
SOURCE human.
GI:6935438

ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 678)
AUTHORS NIH-MGC
TITLE NIH-MGC
JOURNAL National Institutes of Health, Mammalian Gene Collection (MGC)
COMMENT Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgabbs-roma@nih.gov
Tissue Procurement: ATCC
CDNA Library Preparation: Ling Hong/Rubin Laboratory
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
DNA Sequencing by: National Institutes of Health Intramural
Sequencing Center (NISC)
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LNL at:
www.bio.lnl.gov/bdnp/image/image.html
Plate: LCM55 row: B column: 11
Seq primer: M13RP reverse primer (ABI).

FEATURES

source
1. 678
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:2960938"
/clone_lib="NIH_MGC_17"
/clone_type="rhabdomyosarcoma"
/lab_host="DH10B (phage-resistant)"
/note="Organ: muscle; Vector: pORF7; Site 1: EcoRI;
Site 2: XhoI; cDNA made by oligo-dT priming.
Directionally cloned into EcoRI/XhoI sites using the
following 5' adaptor: GGCACGAG(C). Size-selected 500bp
for average insert size 1.8kb. Library constructed by
Ling Hong in the laboratory of Gerald M. Rubin (University
of California, Berkeley) using ZAP-cDNA synthesis kit
(Stratagene) and Superscript II RT (Life Technologies)."

BASE COUNT

161 a 190 c 159 g 168 t

Query Match 42.1%; Score 588.4; DB 10; Length 678;
Best Local Similarity 92.8%; Pred. No. 1.2e-89;
Matches 648; Conservative 0; Mismatches 1; Indels 49; Gaps 1;

645 GCTAATCCCGAGAGCAGATGAGCAGGAGCAAGCTGCAATAGCCAAAGTTAA 704
678 GCTAATCCCGAGAGCAGATGAGCAGGAGCAAGCTGCAATAGCCAAAGTTAA 619
705 AATGTAGTTTACAGCTAGCTATGAGATGCTGCTCTAGTCCAGAAATCATAGGGG 764
618 AATGTAGTTTACAGCTAGCTATGAGATGCTGCTCTAGTCCAGAAATCATAGGGG 559
765 TATGACTGCTCTCCACCTGCTGGGCTGTAGAGCAAGCTGAGTCTCCCACTGGG 824
558 TATGACTGCTCTCCACCTGCTGGGCTGTAGAGCAAGCTGAGTCTCCCACTGGG 499
825 GGCTGTGCCCCCTCCCTGAGCGGTTCCGTGGGAGCCCCCATCTGTGTTCAATAGTGTG 884
498 GGCTGTGCCCCCTCCCTGAGCGGTTCCGTGGGAGCCCCCATCTGTGTTCAATAGTGTG 439
885 AGAATGATGCTAAAGCCCTCTGCTGCTGTCACATCCACAGCAGGGGTGGGGGCT 944
438 AGAATGATGCTAAAGCCCTCTGCTGCTGTCACATCCACAGCAGGGGTGGGGGCT 428
945 GCGTGGGAGCAATCCATCTGAGTGTCTCTCAGCTAGGTGAGCAGGAGACTGGC 1004
427 GCGTGGGAGCAATCCATCTGAGTGTCTCTCAGCTAGGTGAGCAGGAGACTGGC 368
1005 GGGGATGCTCCAGATGTGGGTATCTGTACTGAGGAGGCTATCTGACCTCCCGA 1064
367 GGGGATGCTCCAGATGTGGGTATCTGTACTGAGGAGGCTATCTGACCTCCCGA 308
1065 CAGGGAGACATCCCGAGGCGAGCCGAGGGGTGAGGGGAGAGGTGACACTCAGCATGAG 1124
307 CAGGGAGACATCCCGAGGCGAGCCGAGGGGTGAGGGGAGAGGTGACACTCAGCATGAG 248

QY 1125 CCAAGACTGGGTCAGGAGCAGGTGTGTTTGAAGCCAGGACTGGGGCGGGGTGGGCG 1184
 DB 247 CCAAGACTGGGTCAGGAGCAGGTGTGTTTGAAGCCAGGACTGGGGCGGGGTGGGCG 188
 QY 1185 CGGGGCTTCTTCCCTCATTTGCTTTCATGAAAGCTCAAGACGCCAAACAGGCTT 1244
 DB 187 CGGGGCTTCTTCCCTCATTTGCTTTCATGAAAGCTCAAGACGCCAAACAGGCTT 128
 QY 1245 TCCCTCTCTCGAGTTGATATCCGAATCTTTTGACTCTCTGTGTGTTAATGTT 1304
 DB 127 TCCCTCTCTCGAGTTGATATCCGAATCTTTTGACTCTCTGTGTGTTAATGTT 68
 QY 1305 TATTTTGTAAAAATATAATATAATGTTAATAAAA 1342
 DB 67 TATTTTGTAAAAATATAATATAATGTTAATAAAA 30

RESULT 8
 BG827019 776 bp mRNA linear EST 22-MAY-2001
 LOCUS 602749053F1 NIH_MGC_17 Homo sapiens cDNA clone IMAGE:4901869 5'
 DEFINITION mRNA sequence.
 ACCESSION BG827019 GI:14174606
 VERSION BG827019.1
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 1. (bases 1 to 776)
 NIH-MGC http://imgc.nci.nih.gov/
 National Institutes of Health, Mammalian Gene Collection (MGC)
 Unpublished (1999)
 Contact: Robert Strausberg, Ph.D.
 Email: cgabbs-remail.nih.gov
 Tissue Procurement: ATCC
 CDNA Library Preparation: Ling Hong/Rubin Laboratory
 CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
 DNA Sequencing by: Incyte Genomics, Inc.
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LNL at:
 http://image.llnl.gov
 Plate: LNCM1797 row: j column: 14
 High quality sequence stop: 771.

FEATURES
 source Location/Qualifiers
 1..776
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="IMAGE:4901869"
 /clone_1ib="NIH MGC 17"
 /rissue_type="rhabdomyosarcoma"
 /lab_host="DH10B (phage-resistant)"
 /note="Organ: muscle; Vector: pOTB7; Site: 1: EcoRI;
 Site: 2: XhoI; CDNA made by oligo-dT priming.
 Directionally cloned into EcoRI/XhoI sites using the
 following 5' adaptor: GGCACGAG(G). Size-selected >500bp
 for average insert size 1.8kb. Library constructed by
 Ling Hong in the laboratory of Gerald M. Rubin (University
 of California, Berkeley) using ZAP-CDNA synthesis kit
 (Stratagene) and Superscript II RT (Life Technologies)."
 ORIGIN 158 a 228 c 222 g 168 t

Query Match 41.1%; Score 574.6; DB 12; Length 776;
 Best Local Similarity 97.7%; Pred. No. 2,4e-87;
 Matches 583; Conservative 0; Mismatches 14; Indels 0; Gaps 0;

QY 420 GTTACTCTGCACTGCGCCCGCAGGTGGGCTGTTTCAAGACTGATGAGAG 479
 DB 180 GTTACTCTGCACTGCGCCCGCAGGTGGGCTGTTTCAAGACTGATGAGAG 239
 QY 480 CAAGTGTCCCTGATCCCAAGACCAATATGTGAAGGCTCTGCTGACCTATCTGA 539

DB 240 CAAGTGTCCCTGATCCCAAGACCAATATGTGAAGGCTCTGCTGACCTATCTGA 299
 QY 540 GGGCTGGGTCACACACTGATATCTCAAGACGCTGGGGCTGTCTGTGAAGGAGTACT 599
 DB 300 GGGCTGGGTCACACACTGATATCTCAAGACGCTGGGGCTGTCTGTGAAGGAGTACT 359
 QY 600 TGCACCTGGAGACACTGATGTCACCTGGGAAACCCCTGCAGACAAAGCTAATCCAGAC 659
 DB 360 TGCACCTGGAGACACTGATGTCACCTGGGAAACCCCTGCAGACAAAGCTAATCCAGAC 419
 QY 660 AGACAGATGTGACACAGACAAAGCTGCAATATATGCCAAATGTTAAATGTAGTTTACCA 719
 DB 420 AGACAGATGTGACACAGACAAAGCTGCAATATATGCCAAATGTTAAATGTAGTTTACCA 479
 QY 720 GCTTACTATGGAACCTGCTGCTCTCACTAGTCCAGAGATTCATGGGGATATGACTCTCTCC 779
 DB 480 GCTTACTATGGAACCTGCTGCTCTCACTAGTCCAGAGATTCATGGGGATATGACTCTCTCC 539
 QY 780 AACCTGTGGGCTGTAGGACAGCTCAGGCTAGTCTCCCACTGGGGGCTGTGCCCTCCG 839
 DB 540 AACCTGTGGGCTGTAGGACAGCTCAGGCTAGTCTCCCACTGGGGGCTGTGCCCTCCG 599
 QY 840 TGGGACGGTTCGCTGGGACAGCCCATCACTGTGTTCAATAGTGTGAATGTAGTTAAG 899
 DB 600 TGGGACGGTTCGCTGGGACAGCCCATCACTGTGTTCAATAGTGTGAATGTAGTTAAG 659
 QY 900 CCCCTGCTCTGCTGTGTCACATGTCACAGCAGCGGTGGGGCTGCTGGGACATATC 959
 DB 660 CCCCTGCTCTGCTGTGTCACATGTCACAGCAGCGGTGGGGCTGCTGGGACATATC 719
 QY 960 ATCGTGAATGTCTCTGCTGACCTTAGTGTGACAGAGACTTGGCGGGGATGCTCTCC 1016
 DB 720 ATCGTGAATGTCTCTGCTGACCTTAGTGTGACAGAGACTTGGCGGGGATGCTCTCC 776

RESULT 9
 BE250262 770 bp mRNA linear EST 13-JUL-2000
 LOCUS 600943273F1 NIH_MGC_17 Homo sapiens cDNA clone IMAGE:2959770 5'
 DEFINITION mRNA sequence.
 ACCESSION BE250262
 VERSION BE250262.1 GI:9120370
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 1. (bases 1 to 770)
 NIH-MGC http://imgc.nci.nih.gov/
 National Institutes of Health, Mammalian Gene Collection (MGC)
 Unpublished (1999)
 Contact: Robert Strausberg, Ph.D.
 Email: cgabbs-remail.nih.gov
 Tissue Procurement: ATCC
 CDNA Library Preparation: Ling Hong/Rubin Laboratory
 CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
 DNA Sequencing by: Incyte Genomics, Inc.
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LNL at: image.llnl.gov
 Plate: LNCM52 row: a column: 19
 High quality sequence stop: 730.

FEATURES
 source Location/Qualifiers
 1..770
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="IMAGE:2959770"
 /clone_1ib="NIH MGC 17"
 /rissue_type="rhabdomyosarcoma"
 /lab_host="DH10B (phage-resistant)"
 /note="Organ: muscle; Vector: pOTB7; Site: 1: EcoRI;
 Site: 2: XhoI; CDNA made by oligo-dT priming.
 Directionally cloned into EcoRI/XhoI sites using the

LOCUS BO647652 935 bp mRNA linear EST 15-JUL-2002
 DEFINITION AGENCOURT_8288722 NIH_MGC_100 Homo sapiens cDNA clone IMAGE:6292938
 5', mRNA sequence.
 ACCESSION BO647652
 VERSION BO647652.1 GI:21771824
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 1 (bases 1 to 935)
 NIH-MGC http://mgi.nci.nih.gov/
 National Institutes of Health, Mammalian Gene Collection (MGC)
 Unpublished (1999)
 Contact: Robert Strausberg, Ph.D.
 Email: cgaabs-remail.nih.gov
 Tissue Procurement: CGAP (Stanford)
 cDNA Library Preparation: Rubin Laboratory
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
 DNA Sequencing by: Agencourt Bioscience Corporation
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LNL at:
 http://image.llnl.gov
 Plate: LHC2495 row: c column: 19
 High quality sequence stop: 476.
 Location/Qualifiers
 1. 935
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone_image="6292938"
 /clone_lib="NIH_MGC_100"
 /issue_type="hepatocellular carcinoma, cell line"
 /lab_host="DH10B (phage-resistant)"
 /note="Organ: liver; Vector: pOT7; Site 1: XhoI; Site 2:
 EcoRI. cDNA made by oligo-dT priming. Directionally cloned
 into EcoRI/XhoI sites using the following 5' adaptor:
 GGACGAG(G). Size selected >500bp for average insert size
 1.8kb. Library constructed by Ling Hong in the laboratory
 of Gerald M. Rubin (University of California, Berkeley)
 using ZAP-cDNA synthesis kit (Stratagene) and Superscript
 II RT (Life Technologies). Note: this is a NIH_MGC
 Library."
 BASE COUNT 201 a 253 c 268 g 207 t 6 others
 ORIGIN
 Query Match 39.7%; Score 554; DB 14; Length 935;
 Best Local Similarity 90.1%; Pred. No. 6.3e-84;
 Matches 727; Conservative 0; Mismatches 63; Indels 17; Gaps 12;
 QY 420 GTTACCTCTCAGCTGCCCCCAGGCTGGCAGCTGCTTTTCAAGACTGATGAGC 479
 DB 72 GTACCAACAGACAGCACTGAGAGCTGGCAGCTGCTTTTCAAGACTGATGAGC 131
 QY 480 CAAGTGGTCCCTATCCCAAG-ACCACATATGTGAAGGCTTGGCTGACTTATCTG 538
 DB 132 CAAGTGGTCCCTATCCCAAGAAACCAATATGTGAAGGCTTGGCTGACTTATCTG 191
 QY 539 AGGGCTCGGCTGACGAGTACATCTCTGAGAGCTGGGCTTGGCTGAGGAGTAC 598
 DB 192 AGGGCTCGGCTGACGAGTACATCTCTGAGAGCTGGGCTTGGCTGAGGAGTAC 251
 QY 599 TTGACATGAGCAGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 658
 DB 252 TTGACATGAGCAGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 311
 QY 659 CAGACAGATGTGACAGAGCAAAAGTGAATTAATGTTAAATGTTAAATGTTAAATG 718
 DB 312 CAGACAGATGTGACAGAGCAAAAGTGAATTAATGTTAAATGTTAAATGTTAAATG 371
 QY 719 AGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 777
 DB 372 AGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 431

QY 778 CCACCTGTGGGCTGTATAGCAAGCTAGCTATCTCCAC-AGTGGGGCTGTGCCCCCT 836
 DB 432 CCACCTGTGGGCTGTATAGCAAGCTAGCTATCTCCAC-AGTGGGGCTGTGCCCCCT 491
 QY 837 CCTGGGAGATGTTCCGTGGGAGAGCCCATCAGCTGTGTTCAATATGTGAAATGAGTAA 896
 DB 492 CCTGGGAGATGTTCCGTGGGAGAGCCCATCAGCTGTGTTCAATATGTGAAATGAGTAA 551
 QY 897 AAGCCCTGTGCTGTGCTGACATGACAC-AGCAGGCGGT-GGGGCTGTGCTGGGAGC 954
 DB 552 AAGCCCTGTGCTGTGCTGACATGACAC-AGCAGGCGGT-GGGGCTGTGCTGGGAGC 611
 QY 955 AATCCA-TGTGAGATGTTCTCTACCTTAGTCTTGAGACAGAGAC-TTGGCGGGGAGT 1012
 DB 612 AATCCATGTGAGATGTTCTCTACCTTAGTCTTGAGACAGAGAC-TTGGCGGGGAGT 671
 QY 1013 CTCACAG--ATGTGGTATTTCTGTACTTG-GGAGGCTATCTCTACCTTCCGACAGGG 1069
 DB 672 CTCACAGATGTGGGATGTTCTGTACTTGAGAGGAGGCTATCTCTACCTTCCGACAGGG 731
 QY 1070 GACACTCCAGGCGCAGCC--AGGGGTGAGGGGAGAGGTGACACCTCAGCATGAGCC 1126
 DB 732 GACACTCCAGGCGCAGCC--AGGGGTGAGGGGAGAGGTGACACCTCAGCATGAGCC 791
 QY 1127 --AAGACTGGGGTCAAGG--AGCAGGTGTGTTTGAAGCAGACCTGGGGGGGGTGG 1182
 DB 792 CAAACTTGGGGGTCAAGGAAACAATGTTGTTTGAAGCAGACCTGGGGGGGGTGG 851
 QY 1183 GCCGGGCTTCTGTGCTCATTTGCTT 1209
 DB 852 GTGAGACCCCGGGCTTTTCTTCTTCT 878

RESULT 12
 LOCUS BM692817 548 bp mRNA linear EST 28-FEB-2002
 DEFINITION UI-E-CKI-afk-m-06-0-UI.r1 UI-E-CKI Homo sapiens cDNA clone
 UI-E-CKI-afk-m-06-0-UI 5', mRNA sequence.
 ACCESSION BM692817
 VERSION BM692817.1 GI:19006075
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 1 (bases 1 to 548)
 Bonaldo,M.F., Lennon,G. and Soares,M.B.
 Normalization and subtraction: two approaches to facilitate gene
 discovery
 JOURNAL Genome Res. 6 (9), 791-806 (1996)
 MEDLINE 97044477
 COMMENT
 Contact: Soares, MB
 Program for Rat Gene Discovery and Mapping
 University of Iowa
 451 Eckstein Medical Research Building Iowa City, IA 52242, USA
 Tel: 319 335 8250
 Fax: 319 335 9565
 Email: mscares@blue.weeg.uiowa.edu
 Tissue Procurement: Dr. Gregg Hageman
 cDNA Library preparation: Dr. M. Bento Soares, University of Iowa
 DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
 Clone Distribution: Researchers may obtain clones from Research
 Genetics (www.regen.com).
 Seq primer: M13 Reverse.
 Location/Qualifiers
 1. 548
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone_lib="UI-E-CKI-afk-m-06-0-UI"
 /issue_type="Retina Foveal and Macular"
 /dev_stage="adult"

FEATURES
 source

Db 243 CTCAGCATGAGCCAGAGCTGGGGTCAAGGAGCAGGTGTGTTGAGCCAGACCTGGGGC 184

QY 1174 GGGGGTGGGGCCGGGGCTTTTGTGCTCATTTTGTTCATGAAAGCTCAAGCAGCC 1233

Db 183 GGGGGTGGGGCCGGGGCTTTTGTGCTCATTTTGTTCATGAAAGCTCAAGCAGCC 124

QY 1234 AAACAGAGCTTTTCCCTTCTCGAGTTTGAATATCCAGAACTTTTGTGTTG 1293

Db 123 AAACAGAGCTTTTCCCTTCTCGAGTTTGAATATCCAGAACTTTTGTGTTG 64

QY 1294 GTTAATTTGTTATTTTGTAAAAAATAAATAATTAATTAATAAATGATTTTAC 1353

Db 63 GTTAAATTTGTTATTTTGTAAAAAATAAATAATTAATTAATAAATGATTTTAC 4

QY 1354 AGC 1356

Db 3 AGC 1

RESULT 14

LOCUS B0888182 823 bp mRNA linear EST 16-AUG-2002

DEFINITION AGENCOURT_8615111 NIH_MGC_100 Homo sapiens cDNA clone IMAGE:6299824

ACCESSION B0888182

VERSION B0888182.1 GI:22280196

KEYWORDS EST.

SOURCE human.

ORGANISM Homo sapiens

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

AUTHORS NIH-MGC http://mgs.nci.nih.gov/

TITLE National Institutes of Health, Mammalian Gene Collection (MGC)

JOURNAL Unpublished (1999)

COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgabbs-rc@mail.nih.gov
Tissue Procurement: CGAP (Stanford)
CDNA Library Preparation: Rubin Laboratory
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LNL at:
http://image.lnl.gov
Plate: LCM2513 row: b column: 17
High quality sequence stop: 538.

FEATURES

Source

1. 823

/organism="Homo sapiens"

/db_xref="taxon:9606"

/clone="IMAGE:6299824"

/clone_1b="NIH_MGC_100"

/tissue_type="hepatocellular carcinoma, cell line"

/lab_host="DH10B (phage-resistant)"

/note="Organ: liver; Vector: pOT7; Site 1: XhoI; Site 2: EcoRI; cDNA made by oligo-dT priming. Directionally cloned into EcoRI/XhoI sites using the following 5' adaptor: GGACAGAG(G). Size selected >500bp for average insert size 1.8kb. Library constructed by Ling Hong in the laboratory of Gerald M. Rubin (University of California, Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and Superscript II RT (Life Technologies). Note: this is a NIH_MGC Library."

BASE COUNT 169 a 228 c 256 g 170 t

ORIGIN

Query Match 38.2%; Score 533.4; DB 14; Length 823;

Best Local Similarity 95.3%; Pred. No. 2e-80;

Matches 571; Conservative 0; Mismatches 26; Indels 2; Gaps 2;

QY 420 GTTACTCTTCACTCTGCCCCCAGGTGGACGTGCTTTTTCAGACTGATGAGAC 479

Db 73 GTACACCAACGAGACCACTGAGACAGGTGGACGTGCTTTTTCAGACTGATGAGAC 132

QY 480 CAAGTGTCTCTGATCCCAAGACCATATATGTGAAGGCTTGTGACTATCTGA 539

Db 133 CAAGTGTCTCTGATCCCAAGACCATATATGTGAAGGCTTGTGACTATCTGA 192

QY 540 GGGCTTGGCTGACACAGCTACTATCTCTGACAGCTGGCTTGTGAGGAGTACT 599

Db 193 GGGCTTGGCTGACACAGCTACTATCTCTGACAGCTGGCTTGTGAGGAGTACT 252

QY 600 TGCATGAGAGACATGACATGACCTGGGACCCCTGAGACAAACTAATCCAGAC 659

Db 253 TGCATGAGAGACATGACATGACCTGGGACCCCTGAGACAAACTAATCCAGAC 312

QY 660 AGACAGATGTGACACAGACAAAGCTGCAATTAATGCCAAATGTAAATGTGATTAACA 719

Db 313 AGACAGATGTGACACAGACAAAGCTGCAATTAATGCCAAATGTGATTAACA 372

QY 720 GCTTACCTATGGAGCTGCTGCTCTCTAGTCCAGAAATCATGGGGTATGACTGCTCTC 779

Db 373 GCTTACCTATGGAGCTGCTGCTCTCTAGTCCAGAAATCATGGGGTATGACTGCTCTC 432

QY 780 AACCTGTGGGCTGTAGAGAGCTGAGTCTCCAGAGGCTGAGGCTGAGGCTGCTCC 839

Db 433 AACCTGTGGGCTGTAGAGAGCTGAGTCTCCAGAGGCTGAGGCTGAGGCTGCTCC 492

QY 840 TGGAGCGTTCGCTGGGAGACCCCATCATCTGTGTTCAATAGTGTGAGAAATGTAGTAAAG 899

Db 493 TGGAGCGTTCGCTGGGAGACCCCATCATCTGTGTTCAATAGTGTGAGAAATGTAGTAAAG 552

QY 900 CCCCTGCTCTGCTGTGTCATGACATGACAGACAGCGGTGGGGCTGCTGGGGCAATTC 959

Db 553 CCCCTGCTCTGCTGTGTCATGACATGACAGACAGCGGTGGGGCTGCTGGGGCAATTC 612

QY 960 ATC-GTGGAGTGTCTCTGACG-TTATGCTGTGACAGAGACTTGGCGGGGATGCTCC 1016

Db 613 ATCGGTGAGTGTCTCTGACGTTTAATGTGACAGAGAGACTGCGGGGGGATGC 671

RESULT 15

LOCUS A1669816 529 bp mRNA linear EST 15-DEC-1999

DEFINITION LUI3108.X1 NCI CGAP_P128 Homo sapiens cDNA clone IMAGE:2252679 3'

ACCESSION A1669816

VERSION A1669816.1 GI:4834590

KEYWORDS EST.

SOURCE human.

ORGANISM Homo sapiens

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap/

TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index

JOURNAL Unpublished (1997)

COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgabbs-rc@mail.nih.gov
Tissue Procurement: Michael J. Brownstein, M.D., Ph.D., Michael R. Emmert-Buck, M.D., Ph.D.
CDNA Library Preparation: M. Bento Soares, Ph.D.
CDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LNL at:
www-bio.lnl.gov/btrp/image/image.html
Insert Length: 912 Std Error: 0.00
Seq primer: -40UP from Glibco
High quality sequence stop: 455.

FEATURES

Source

1. 529

/organism="Homo sapiens"

/db_xref="taxon:9606"

/clone="IMAGE:2252679"

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 29, 2003, 04:56:03 ; Search time 2476 Seconds

(without alignments)
16420.283 Million cell updates/sec

Title: US-09-989-919-15

Perfect score: 1397
Sequence: 1 ggtcgtcaccgtctacgga.....aaaaaaaaaaagcggtc 1397

Scoring table: OLIGO_NUC
Gapop 60.0 , Gapext 60.0

Searched: 2054640 seqs, 14551402878 residues

Word size : 0

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 1000 summaries

Database :

GenBankl:*
1: gb_ba:*
2: gb_hcg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pac:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_scs:*
12: gb_sy:*
13: gb_un:*
14: gb_vl:*
15: em_ba:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_mu:*
20: em_om:*
21: em_or:*
22: em_ov:*
23: em_pat:*
24: em_ph:*
25: em_pl:*
26: em_ro:*
27: em_scs:*
28: em_un:*
29: em_vl:*
30: em_hcg_hum:*
31: em_hcg_inv:*
32: em_hcg_other:*
33: em_hcg_mus:*
34: em_hcg_pin:*
35: em_hcg_rtd:*
36: em_hcg_mam:*
37: em_hcg_vrt:*
38: em_sy:*
39: em_hcgo_hum:*
40: em_hcgo_mus:*
41: em_hcgo_other:*

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1397	100.0	1397	6 AX480725	AX480725 Sequence
2	1258	90.1	128466	2 AC013480	AC013480 Homo sapi
3	876	62.7	2461	6 AX056460	AX056460 Sequence
4	867	62.1	1493	6 BC007901	BC007901 Homo sapi
5	442	31.6	470	6 AX480724	AX480724 Sequence
6	322	23.0	427	6 AX408491	AX408491 Sequence
7	152	10.9	462	11 G21939	G21939 humanSTS W
8	43	3.1	94366	2 AC112468	AC112468 Rattus no
9	43	3.1	176771	2 AC122406	AC122406 Mus muscu
10	36	2.6	1586	10 BC022157	BC022157 Mus muscu
11	36	2.6	1886	6 AX056461	AX056461 Sequence
12	33	2.4	69492	2 AC111157	AC111157 Homo sapi
13	32	2.3	164796	9 AC091069	AC091069 Homo sapi
14	32	2.3	165201	2 AC011963	AC011963 Homo sapi
15	30	2.1	1416	9 HSA420562	AX376502 Sequence
16	30	2.1	2457	6 AX376502	AX376502 Sequence
17	30	2.1	3278	9 HSM800650	AL080138 Homo sapi
18	30	2.1	111723	2 AC020935	AC020935 Homo sapi
19	30	2.1	127992	2 AC105036	AC105036 Homo sapi
20	30	2.1	148263	9 AC091988	AC091988 Homo sapi
21	30	2.1	186248	2 AC095218	AC095218 Rattus no
22	29	2.1	942	9 HSA406937	AJ406937 Homo sapi
23	29	2.1	1217	8 DCU38483	U38483 D. dentatus ca
24	29	2.1	1229	10 BC004689	BC004689 Mus muscu
25	29	2.1	3329	6 AX407231	AX407231 Sequence
26	29	2.1	4301	6 AX480779	AX480779 Sequence
27	29	2.1	82360	9 AF240580	AF240580 Homo sapi
28	29	2.1	143688	2 AC127169	AC127169 Medicago
29	29	2.1	145329	2 AC055850	AC055850 Homo sapi
30	29	2.1	157871	2 AC025074	AC025074 Homo sapi
31	29	2.1	167603	2 AC068261	AC068261 Homo sapi
32	29	2.1	170355	2 AC111904	AC111904 Rattus no
33	29	2.1	173029	2 AC117646	AC117646 Mus muscu
34	29	2.1	190584	2 AC103589	AC103589 Homo sapi
35	29	2.1	198231	2 AC130141	AC130141 Rattus no
36	29	2.1	203474	2 AC069307	AC069307 Homo sapi
37	28	2.0	600	8 ATY13174	Y13174 Arabidopsis
38	28	2.0	839	4 AF043534	AF043534 Arabidopsis
39	28	2.0	1000	9 BC011886	BC011886 Homo sapi
40	28	2.0	1141	3 AF094810	AF094810 Drosophila
41	28	2.0	1209	9 AF087873	AF087873 Homo sapi
42	28	2.0	1318	9 BC034354	BC034354 Homo sapi
43	28	2.0	1439	9 BC000550	BC000550 Homo sapi
44	28	2.0	1560	9 AK026057	AK026057 Homo sapi
45	28	2.0	1569	9 AF131844	AF131844 Homo sapi
46	28	2.0	2405	4 BTACTPRA	X57084 Bovine mRNA
47	28	2.0	2466	10 AF001157	AF001157 Mesocricet
48	28	2.0	2641	3 AB073673	AB073673 Bombyx mori
49	28	2.0	2796	9 AK074352	AK074352 Homo sapi
50	28	2.0	16170	6 AX346171	AX346171 Sequence
51	28	2.0	29501	9 AC010248	AC010248 Homo sapi
52	28	2.0	37661	6 AX147034	AX147034 Sequence
53	28	2.0	46535	2 AC119222	AC119222 Mus muscu
54	28	2.0	51739	4 AC095023	AC095023 Sus scrofa
55	28	2.0	55470	2 AC116989	AC116989 Dictyostere
56	28	2.0	70121	2 AC123693	AC123693 Mus muscu
57	28	2.0	70135	2 AC095482	AC095482 Rattus no
58	28	2.0	70394	2 AC023011	AC023011 Homo sapi
59	28	2.0	71192	2 AC095584	AC095584 Rattus no
60	28	2.0	74188	2 AC021315	AC021315 Homo sapi
61	28	2.0	75002	2 AC023384	AC023384 Homo sapi
62	28	2.0	81276	2 HS276A23	AL449214 Homo sapi
63	28	2.0	92705	9 AC006327	AC006327 Homo sapi
64	28	2.0	98778	2 AC002408	AC002408 Homo sapi
65	28	2.0	102116	2 AC122681	AC122681 Rattus no

Pred. No. is the number of results predicted by chance to have a

[illegible]

```
/rpt_family="(T)n"
repeat_region 3085..3392
/rpt_family="Alu"
repeat_region 3530..3643
/rpt_family="CR1"
repeat_region 3644..3717
/rpt_family="Mariner"
repeat_region 3718..4183
/rpt_family="CR1"
repeat_region 6123..6293
/rpt_family="MIR"
repeat_region 6845..6881
/rpt_family="AT_rich"
repeat_region 6911..6963
/rpt_family="AT_rich"
repeat_region 7910..8008
/rpt_family="AT_rich"
repeat_region 8009..8299
/rpt_family="MIR"
repeat_region 8300..8326
/rpt_family="MIR"
repeat_region 8337..8599
/rpt_family="Alu"
repeat_region 8600..8668
/rpt_family="MIR"
repeat_region 9463..9553
/rpt_family="(TTCA)n"
repeat_region 9890..9923
/rpt_family="AT_rich"
repeat_region 10066..10473
/rpt_family="L2"
repeat_region 10474..10520
/rpt_family="(CA)n"
repeat_region 10521..10746
/rpt_family="L2"
repeat_region 12200..12385
/rpt_family="MIR"
repeat_region 12452..12489
/rpt_family="(CACG)n"
repeat_region 13410..13773
/rpt_family="MALR"
repeat_region 14076..14317
/rpt_family="MIR"
repeat_region 15470..15592
/rpt_family="MIR"
repeat_region 15599..15967
/rpt_family="MALR"
misc_feature 16227..16617
/ncore="similar to EST AA913068 (NID:G3052460) o134e06.s1"
repeat_region 16517..16693
/rpt_family="L1"
repeat_region 16717..16788
/rpt_family="L1"
repeat_region 16988..17009
/rpt_family="AT_rich"
repeat_region 18365..18509
/rpt_family="MER103"
repeat_region 18594..18644
/rpt_family="(TG)n"
repeat_region 19032..19193
/rpt_family="MER1_type"
repeat_region 19359..19668
/rpt_family="L2"
repeat_region 22454..22498
/rpt_family="GA-rich"
repeat_region 25334..25644
/rpt_family="Alu"
repeat_region 25647..25768
/rpt_family="MIR"
repeat_region 25770..25888
/rpt_family="L2"
repeat_region 25893..26193
/rpt_family="Alu"
```

```
misc_feature 25910..26358
/ncore="similar to EST AA578847 (NID:G2357031) nk66a09.s1"
repeat_region 26368..26432
/rpt_family="MIR"
repeat_region 26502..26568
/rpt_family="MIR"
repeat_region 26725..27274
/rpt_family="MALR"
repeat_region 27275..27400
/rpt_family="MIR"
repeat_region 28344..28444
/rpt_family="MIR"
repeat_region 28445..28737
/rpt_family="Alu"
repeat_region 28738..28877
/rpt_family="MIR"
repeat_region 29303..29383
/rpt_family="Alu"
repeat_region 31059..31158
/rpt_family="L1"
repeat_region 31180..31624
/rpt_family="L1"
repeat_region 31626..31773
/rpt_family="Alu"
repeat_region 31774..31993
/rpt_family="Alu"

Query Match 90.1%; Score 1258; DB 9; Length 128466;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 1108; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 62 GAGTACCATGTATCCAGACAGACACATCCCGGAGAAAGTACCGCTGCTGCATCC 121
DB 113798 GAGTACCATGTATCCAGACAGACACATCCCGGAGAAAGTACCGCTGCTGCATCC 113857
QY 122 TACCAACACGAGGAGCTGCTCTCTTCACTGTTCACTGCTGAGGCTGTGATCTGT 181
DB 113858 TACCAACACGAGGAGCTGCTCTCTTCACTGTTCACTGCTGAGGCTGTGATCTGT 113917
QY 182 GAGAGCATGCCCATGTGCGGCTTTGTGTGACCAACCAACCACTGACAGTAG 241
DB 113918 GAGAGCATGCCCATGTGCGGCTTTGTGTGACCAACCAACCACTGACAGTAG 113977
QY 242 CCAGTGGAGAGAGCCCTTCAAGGAGATGAGAGACCTCTGTGAGGTGATAGT 301
DB 113978 CCAGTGGAGAGAGCCCTTCAAGGAGATGAGAGACCTCTGTGAGGTGATAGT 114037
QY 302 GATCCCCATCGGAAGTCAGAGGGGCTGTGAGGTGATGAGAGAGTATCGTCTT 361
DB 114038 GATCCCCATCGGAAGTCAGAGGGGCTGTGAGGTGATGAGAGAGTATCGTCTT 114097
QY 362 CAAGGAGTCAATTAAGGAGAAATGCTTGGCTCCAGAAAGAAATCCAGGCCCTGT 421
DB 114098 CAAGGAGTCAATTAAGGAGAAATGCTTGGCTCCAGAAAGAAATCCAGGCCCTGT 114157
QY 422 TACCTTCACTCTGCCCCCAGAGTCGAGCTGTGCTTTTCAAGACTGATGAGCA 481
DB 114158 TACCTTCACTCTGCCCCCAGAGTCGAGCTGTGCTTTTCAAGACTGATGAGCA 114217
QY 482 AGTGTCCCTGATCCCAACAACAACATATGTGAAGGCTTGGCTGATCTATCTGAG 541
DB 114218 AGTGTCCCTGATCCCAACAACAACATATGTGAAGGCTTGGCTGATCTATCTGAG 114277
QY 542 GCTGGGCTACCAAGCTGATATCTTCAGAGCTGGGCTTGGCTGGAGGAGTACTTG 601
DB 114278 GCTGGGCTACCAAGCTGATATCTTCAGAGCTGGGCTTGGCTGGAGGAGTACTTG 114337
QY 602 CACTGGACAGCATGATGCTACCTGGGAACCCCTGACAGCAAAAGTAAATCCACAGAG 661
DB 114338 CACTGGACAGCATGATGCTACCTGGGAACCCCTGACAGCAAAAGTAAATCCACAGAG 114397
QY 662 ACAGATGTGACCAAGCAAAAGTCAATATATGCAATTTAAATGTGATTTACAGC 721
DB 114397 ACAGATGTGACCAAGCAAAAGTCAATATATGCAATTTAAATGTGATTTACAGC 114457
```


SOURCE Homo sapiens.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 REFERENCE 1 (bases 1 to 1493)
 AUTHORS Strausberg, R.
 TITLE Direct Submission
 JOURNAL Submitted (11-MAY-2001) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA
 REMARK NIH-MGC Project URL: <http://mgc.nci.nih.gov>
 COMMENT Contact: MGC help desk
 Email: cgabs-remail.nih.gov
 Tissue Procurement: ATCC
 cDNA Library Preparation: Rubin Laboratory
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
 DNA Sequencing by: National Institutes of Health Intramural Sequencing Center (NISC)
 Gaithersburg, Maryland
 Web site: <http://www.nisc.nih.gov/>
 Contact: nisc.mgc@nih.gov
 Shevchenko, Y., Wetherby, R.D., Beckstrom-Sternberg, S.M., Benjamin, B., Blakeley, R.M., Bouffard, G.G., Brinkley, C., Brooks, S., Dietrich, N.L., Guan, X., Gupta, J., Ho, S.-L., Karlins, E., Legaspi, R., Lim, M., Maduro, O.L., Masello, C., Mastrian, S.D., McCloskey, J.C., McDowell, J., Pearson, R., Snyder, B., Stantrop, S., Thomas, P.J., Tjongsom, E.B., Touchman, J.W., Tsurgan, C., Vogt, J.L., Walker, M.A., Zhang, L.-H. and Green, E.D.
 Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LNL at: <http://image.lnl.gov>
 Series: IRAL Plate: 20 Row: 3 Column: 15
 This clone was selected for full length sequencing because it passed the following selection criteria: Hexamer frequency OR analysis.
 FEATURES
 source location/Qualifiers
 1. 1493
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="IMAGE:4139786"
 /tissue_type="Muscle, rhabdomyosarcoma"
 /clone_lib="NIH MGC_17"
 /lab_host="DH10B-R"
 /note="Vector: pOT87"
 <1. 625
 /codon_start=2
 /product="unknown (protein for IMAGE:4139786)"
 /protein_id="AA07901.1"
 /db_xref="GI:14043928"
 /translation="ARGELKVTLDLADARVETPCAGSTDCILEFPAPRNTLPSSAQMGCEGNKRRLYNARFFFTYLPASAPSPRLPDLISIVNATGLANGVETLALQLEKVLHLYRSGOYLONSTASSTYOCIPDSITPOBYRCMPSYHSGSLSVENLAEAVDVCEHACOCRAVYVNOTTWTGROLVFPFTGMSQVVPDPPNKTYKASG"
 CDS
 BASE COUNT 360 a 401 c 422 g 310 t
 ORIGIN
 Query Match 62.1%; Score 867; DB 9; Length 1493;
 Best Local Similarity 99.9%; Pred. No. 0;
 Matches 917; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
 Db 537 CAGGTGGGAGCTGCTTTTCAAGCTGATGGAGGAGGAGGCTGCTGATCCCAACA 501
 537 CAGGTGGGAGCTGCTTTTCAAGCTGATGGAGGAGGAGGCTGCTGATCCCAACA 596
 QY 442 CAGGTGGGAGCTGCTTTTCAAGCTGATGGAGGAGGAGGCTGCTGATCCCAACA 501
 Db 502 AGACCAATATGTAAGGCTGCTGATGACTATCTGAGGAGCTGCTGATGACGACTACT 561
 597 AGACCAATATGTAAGGCTGCTGATGACTATCTGAGGAGCTGCTGATGACGACTACT 656
 QY 562 ATCTTGAGAGCTGCTGCTGCTGATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 621
 Db 657 ATCTTGAGAGCTGCTGCTGCTGATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 716

QY 622 ACTGGGAACCCCTGACAGCAAGGTAACATCCAGAGAGAGATGTGACGAGCAAA 681
 Db 717 ACTGGGAACCCCTGACAGCAAGGTAACATCCAGAGAGAGATGTGACGAGCAAA 776
 QY 682 CCGTGAATATATCCCAATGTTTAAATGTGAGTTTACAGCTTACGATATGAGGAGTGTGGC 741
 Db 777 CCGTGAATATATCCCAATGTTTAAATGTGAGTTTACAGCTTACGATATGAGGAGTGTGGC 836
 QY 742 TCCCTGTCAGGAATATATGAGGAGTATGAGTGTGCTCCCAACCTGAGGAGTGTGAGCAAG 801
 Db 837 TCCCTGTCAGGAATATATGAGGAGTATGAGTGTGCTCCCAACCTGAGGAGTGTGAGCAAG 896
 QY 802 CTCAGGAGTATGCTCCCACTGAGGAGGAGTGTGCTCCCTGAGGAGGAGTGTGAGGAGGAGC 861
 Db 897 CTCAGGAGTATGCTCCCACTGAGGAGGAGTGTGCTCCCTGAGGAGGAGTGTGAGGAGGAGC 956
 QY 862 CCATCACTGTGTTCAATATGAGTGTGAGATATGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTG 921
 Db 957 CCATCACTGTGTTCAATATGAGTGTGAGATATGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTG 1016
 QY 922 TCCCAAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 981
 Db 1017 TCCCAAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1076
 QY 982 TAGGTCTGAGCAGAGAGACTTGGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1041
 Db 1077 TAGGTCTGAGCAGAGAGACTTGGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1136
 QY 1042 GGAGGCTATCTGTGACCTCCGACAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1101
 Db 1137 GGAGGCTATCTGTGACCTCCGACAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1196
 QY 1102 AAGAGTGCACACTGACAGATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1161
 Db 1197 AAGAGTGCACACTGACAGATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1256
 QY 1162 AGACCTGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1221
 Db 1257 AGACCTGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1316
 QY 1222 TCAAGACAGCAAAACAGGCTTTTCCCTTCTGAGATTGAATATCCAGAAATCTTTTG 1281
 Db 1317 TCAAGACAGCAAAACAGGCTTTTCCCTTCTGAGATTGAATATCCAGAAATCTTTTG 1376
 QY 1282 TACTTCTTGTGTTAAATTTGTTTATTTTGTAAAAAATAAATAAATAGTATATAA 1341
 Db 1377 TACTTCTTGTGTTAAATTTGTTTATTTTGTAAAAAATAAATAAATAGTATATAA 1436
 QY 1342 ATGATGTTTACAGCAAA 1359
 Db 1437 ATGATGTTTACAGCAAA 1454
 RESULT 5
 AX480724 470 bp DNA linear PAT 12-AUG-2002
 DEFINITION Sequence 14 from Patent WO0242460.
 ACCESSION AX480724
 VERSION AX480724.1 GI:22217466
 KEYWORDS
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 REFERENCE 1
 AUTHORS Macina, R.A., Recipon, H., Pluta, J., Ghosh, M.G., Sun, Y. and Liu, C.
 TITLE Compositions and methods relating to colon specific genes and proteins
 JOURNAL Patent: WO 0242460-A 14 30-MAY-2002;
 Diadexus, Inc. (US)
 FEATURES
 source location/Qualifiers
 1. 470
 /organism="Homo sapiens"

BASE COUNT 98 a 122 c 147 g 103 t
 ORIGIN
 Query Match 31.6%; Score 442; DB 6; Length 470;
 Best Local Similarity 100.0%; Pred. No. 4,2e-249;
 Matches 442; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

572 GCTGGGCTTGGCTGGAGGAGTGAAGTTCAGTGGGAGCACTGATGTCACCTGGGAGC 631
 7 GCTGGGCTTGGCTGGAGGAGTGAAGTTCAGTGGGAGCACTGATGTCACCTGGGAGC 66
 632 CCTGAGACAAAGCTTAACATCCAGACAGAGATGTGACAGACAAAGCTGCAATTA 126
 67 CCTGAGACAAAGCTTAACATCCAGACAGAGATGTGACAGACAAAGCTGCAATTA 126
 692 TGGCAATGTTAAATGATGTTTACAGCTAGCTATGAGATGCTGGCTCTTACTGCA 751
 127 TGGCAATGTTAAATGATGTTTACAGCTAGCTATGAGATGCTGGCTCTTACTGCA 186
 752 GGAATCATGGGGGCTATGACTGCTCTCCAACTGCTGGGCTGTAGAGAACTCAGGCTAG 811
 187 GGAATCATGGGGGCTATGACTGCTCTCCAACTGCTGGGCTGTAGAGAACTCAGGCTAG 246
 812 TCTCCCACTGGGGGCTGTGCTCCCTCCCTGGAAGGCTTCCTGGGAGCCCATCACTGT 871
 247 TCTCCCACTGGGGGCTGTGCTCCCTCCCTGGAAGGCTTCCTGGGAGCCCATCACTGT 306
 872 GTTCATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 931
 307 GTTCATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 366
 932 GGGGCTGGGGGCTGGCTGGGAGCAATTCATGAGATGTTCTCTCAGCTTAGCTGGA 991
 367 GGGGCTGGGGGCTGGCTGGGAGCAATTCATGAGATGTTCTCTCAGCTTAGCTGGA 426
 992 CAGGAGACTTGGGGGGGATGC 1013
 427 CAGGAGACTTGGGGGGGATGC 448

RESULT 6
 AX408491/c 427 bp DNA linear PAT 14-JUN-2002
 LOCUS AX408491
 DEFINITION Sequence 1138 from Patent WO0229103.
 ACCESSION AX408491
 VERSION AX408491.1 GI:21441196
 KEYWORDS
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE
 1 Alvarez, C., Horne, D., Peres-da-Silva, S. and Vockley, J. G.
 Gene expression profiles in liver cancer
 Patent: WO 0229103-A 1138 11-APR-2002;
 JOURNAL GENE LOGIC INC (US)
 FEATURES
 source location/Qualifiers
 1..427
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /note="EMBL/GenBank Accession No. AA451877"

BASE COUNT 107 a 126 c 92 g 102 t
 ORIGIN
 Query Match 23.0%; Score 322; DB 6; Length 427;
 Best Local Similarity 99.5%; Pred. No. 3e-178;
 Matches 422; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

936 GTGGGGGCTGGCTGGGAGCAATCATGCTGAGATGTTCTCTCAGCTTAGCTGGACAG 995
 427 GTGGGGGCTGGCTGGGAGCAATCATGCTGAGATGTTCTCTCAGCTTAGCTGGACAG 368

996 AGACTTGGCGGGGAGTCTCCAGAGATGCGGTGATTTCTTACTGGGAGGCTATCTG 1055
 367 AGACTTGGCGGGGAGTCTCCAGAGATGCGGTGATTTCTTACTGGGAGGCTATCTG 308
 1056 ACTTCCCGACAGGGGACACTCCAGGCGAGCCAGGGGCTCAGGGGAGAGTCCACACT 1115
 307 ACTTCCCGACAGGGGACACTCCAGGCGAGCCAGGGGCTCAGGGGAGAGTCCACACT 248
 1116 CAGCATGAGCCCAAGACTGGGGGTTCAGGGAGCAGGTGTGTTGAGCCAGGACCTGGGCG 1175
 247 CAGCATGAGCCCAAGACTGGGGGTTCAGGGAGCAGGTGTGTTGAGCCAGGACCTGGGCG 188
 1176 GGGTGGGCGGGGCGCTTTCTGCTCTCATTTGCTTTCAATGAAGCTCAAGAGCCAA 1235
 187 GGGTGGGCGGGGCGCTTTCTGCTCTCATTTGCTTTCAATGAAGCTCAAGAGCCAA 128
 1236 ACCAGGCTTTCCCTCTCTCGAGTTGAATATCCGAATCTTTTACTCTGTTGGT 1295
 127 ACCAGGCTTTCCCTCTCTCGAGTTGAATATCCGAATCTTTTACTCTGTTGGT 68
 1296 TAAATGTTTAAATTTTAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1355
 67 TAAATGTTTAAATTTTAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 8
 1356 CAAA 1359
 7 CAAA 4

RESULT 7
 G21939/c 462 bp DNA linear STS 31-MAY-1996
 LOCUS G21939
 DEFINITION human STS WI-15430, sequence tagged site.
 ACCESSION G21939
 VERSION G21939.1 GI:1342265
 KEYWORDS STS; STS sequence; primer; sequence tagged site.
 SOURCE Homo sapiens STS derived from sequences in dbEST and the Unigene collection.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE
 1 Hudson, T.
 Whitehead Institute/MIT Center for Genome Research; Physically
 Mapped STS
 Unpublished (1995)

JOURNAL
 COMMENT
 Contact: Thomas Hudson
 Whitehead Institute/MIT Center for Genome Research
 Whitehead Institute for Biomedical Research
 9 Cambridge Center, Cambridge MA 02142 USA
 Tel: 617 252 1900
 Fax: 617 252 1902
 Email: thudson@genome.wi.mit.edu

Primer A: AGGAGAGTTGCTGTA
 Primer B: GCTCAAGACAGCCAAAC
 STS size: 150
 PCR Profile:
 Presoak:
 Denaturation:
 Annealing: 56 degrees C
 Polymerization:
 PCR Cycles: 35
 Thermal Cycler:
 Protocol:
 Template: 10 ng
 Primer: each 5 pm
 dNTPs: each 4 mM
 Tag Polymerase: 0.025 units/ul
 Total Vol: 20 ul
 Buffer:


```

* 9745 9844: gap of unknown length
* 9845 11023: contig of 1179 bp in length
* 11024 11123: gap of unknown length
* 11124 12519: contig of 1396 bp in length
* 12520 12619: gap of unknown length
* 12620 13673: contig of 1054 bp in length
* 13674 13773: gap of unknown length
* 13774 15620: contig of 1847 bp in length
* 15621 15720: gap of unknown length
* 15721 17466: contig of 1746 bp in length
* 17467 17566: gap of unknown length
* 17567 19136: contig of 1570 bp in length
* 19137 19236: gap of unknown length
* 19237 20973: contig of 1737 bp in length
* 20974 21073: gap of unknown length
* 21074 22091: contig of 1018 bp in length
* 22092 22191: gap of unknown length
* 22192 23778: contig of 1587 bp in length
* 23779 23878: gap of unknown length
* 23879 25293: contig of 1415 bp in length
* 25294 25983: gap of unknown length
* 25984 27083: gap of unknown length
* 27084 29408: contig of 2325 bp in length
* 29409 29508: gap of unknown length
* 29509 31243: contig of 1735 bp in length
* 31244 31343: gap of unknown length
* 31344 33604: contig of 2261 bp in length
* 33605 33704: gap of unknown length
* 33705 35038: contig of 1334 bp in length
* 35039 35138: gap of unknown length
* 35139 36399: contig of 1261 bp in length
* 36400 37623: contig of 1124 bp in length
* 37624 37723: gap of unknown length
* 37724 40199: contig of 2376 bp in length
* 40100 40199: gap of unknown length
* 40200 42674: contig of 2475 bp in length
* 42675 42774: gap of unknown length
* 42775 45617: contig of 2843 bp in length
* 45618 45717: gap of unknown length
* 45718 47425: contig of 1708 bp in length
* 47426 47525: gap of unknown length
* 47526 48809: contig of 1284 bp in length
* 48810 48909: gap of unknown length
* 48910 50692: contig of 1783 bp in length
* 50693 50792: gap of unknown length
* 50793 52171: contig of 1379 bp in length
* 52172 52271: gap of unknown length
* 52272 53845: contig of 1574 bp in length
* 53846 53945: gap of unknown length
* 53946 56501: contig of 2556 bp in length
* 56502 56601: gap of unknown length
* 56602 59622: contig of 3021 bp in length
* 59623 59722: gap of unknown length
* 59723 62615: contig of 2893 bp in length
* 62616 62715: gap of unknown length
* 62716 65623: contig of 2908 bp in length
* 65624 65723: gap of unknown length
* 65724 67513: contig of 1790 bp in length
* 67514 70311: contig of 2698 bp in length
* 70312 70411: gap of unknown length
* 70412 72769: contig of 2358 bp in length
* 72770 72869: gap of unknown length
* 72870 76193: contig of 3324 bp in length
* 76194 76293: gap of unknown length
* 76294 79434: contig of 3141 bp in length
* 79435 79534: gap of unknown length
* 79535 84177: contig of 4643 bp in length
* 84178 84277: gap of unknown length
* 84278 88578: contig of 4301 bp in length
* 88579 88678: gap of unknown length

```

```

FEATURES             * 88679 94366: contig of 5688 bp in length.
                        Location/Qualifiers
                        source
                        /organism="Rattus norvegicus"
                        /db_xref="taxon:10116"
                        /clone="CH250-12986"
BASE COUNT      22805 a 22369 c 21709 g 21135 t 4348 others
ORIGIN
Query Match          3.1% Score 43; DB 2; Length 94366;
Best Local Similarity 100.0%; Pred. No. 1.8e-13;
Matches 43; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY      203 GCCTTGTGTCACACACACACACCTTGACAGGTGACCCAG 245
Db      24116 GCCTTGTGTCACACACACACACCTTGACAGGTGACCCAG 24158

RESULT 9
AC122406/c
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT
----- Genome Center -----
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: http://genome.wustl.edu/gsc/index.shtml
Contact: submissions@watson.wustl.edu
Project information
Center project name: M_BB0126114
----- Summary Statistics -----
Sequencing vector: MJ3; 0%
Sequencing vector: plasmid; 100%
Chemistry: Dye-terminator Big Dye; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.990319
Consensus quality: 171598 bases at least Q40
Consensus quality: 172211 bases at least Q30
Consensus quality: 172737 bases at least Q20
Insert size: 186000; agarose-fp
Quality coverage: 11.87 in Q20 bases; agarose-fp
Quality coverage: 11.02 in Q20 bases; sum-of-contigs
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 12 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.

```

* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved

1 57: contig of 57 bp in length
58 157: gap of unknown length
158 3050: contig of 2893 bp in length
3051 3150: gap of unknown length
3151 5182: contig of 2032 bp in length
5183 5283: gap of unknown length
5283 14170: contig of 888 bp in length
14171 14270: gap of unknown length
14271 22115: contig of 7845 bp in length
22116 22215: gap of unknown length
22216 32755: contig of 10539 bp in length
32756 32855: gap of unknown length
32856 46558: contig of 13704 bp in length
46559 46659: gap of unknown length
46660 65170: contig of 18512 bp in length
65171 65270: gap of unknown length
65271 83892: contig of 18622 bp in length
83893 83992: gap of unknown length
83993 103925: contig of 19933 bp in length
103926 104025: gap of unknown length
104026 133284: contig of 29259 bp in length
133285 133385: gap of unknown length
133386 176771: contig of 43387 bp in length.

Location/Qualifiers
1. 176771
/organism="Mus musculus"
/db_xref="taxon:10090"
/chromosome="UNK"
/clone="RP24-126L14"
1. 57
/note="assembly_name:Contig11"
158. 3050
/note="assembly_name:Contig15"
3151. 5182
/note="assembly_name:Contig16"
5283. 14170
/note="assembly_name:Contig17"
14271. 22115
/note="assembly_name:Contig18"
22216. 32754
/note="assembly_name:Contig19"
32855. 46558
/note="assembly_name:Contig20"
46659. 65170
/note="assembly_name:Contig21"
65271. 83892
/note="assembly_name:Contig22"
83993. 103925
/note="assembly_name:Contig23"
104026. 133284
/note="assembly_name:Contig24"
133385. 176771
/note="assembly_name:Contig25"

FEATURES
source
misc_feature
1. 57
/note="assembly_name:Contig11"
misc_feature
158. 3050
/note="assembly_name:Contig15"
misc_feature
3151. 5182
/note="assembly_name:Contig16"
misc_feature
5283. 14170
/note="assembly_name:Contig17"
misc_feature
14271. 22115
/note="assembly_name:Contig18"
misc_feature
22216. 32754
/note="assembly_name:Contig19"
misc_feature
32855. 46558
/note="assembly_name:Contig20"
misc_feature
46659. 65170
/note="assembly_name:Contig21"
misc_feature
65271. 83892
/note="assembly_name:Contig22"
misc_feature
83993. 103925
/note="assembly_name:Contig23"
misc_feature
104026. 133284
/note="assembly_name:Contig24"
misc_feature
133385. 176771
/note="assembly_name:Contig25"

BASE COUNT 45173 a 42065 c 42632 g 45785 t 1116 others
ORIGIN

Query Match 3.1%; Score 43; DB 2; Length 176771;
Best Local Similarity 100.0%; Pred. No. 1.8e-13;
Matches 43; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 203 GCCTTTGTGTCACCAACGACGACCTGAGCGTGAGCCAG 245
|||||
DB 57080 GCCTTTGTGTCACCAACGACGACCTGAGCGTGAGCCAG 57038
|||||

RESULT 10
LOCUS BC022157 1586 bp mRNA linear ROD 07-AUG-2002
DEFINITION Mus musculus, clone IMAGE:5134400, mRNA, partial cds.
ACCESSION BC022157

VERSION BC022157.1 GI:18381057
KEYWORDS house mouse.
SOURCE Mus musculus.
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 1586)
REFERENCE Strausberg, R.
TITLE Direct Substitution
JOURNAL Submitted (24-JAN-2002) National Institutes of Health, Mammalian
Gene Collection (MGC), Cancer Genomics Office, National Cancer
Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
USA

REMARK NIH-MGC Project URL: <http://mgc.nci.nih.gov>.
COMMENT Contact: MGC help desk
Email: cgabbs-r@mail.nih.gov
Tissue Procurement: Jeffrey E. Green, M.D.
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
DNA Sequencing by: Genome Sequence Centre,
BC Cancer Agency, Vancouver, BC, Canada
info@cgsc.bc.ca
Steven Jones, Jennifer Asano, Ian Bosdet, Yaron Butterfield,
Susanna Chan, Readman Chiu, Chris Fjell, Erin Garland, Ram Guin,
Leticia Hsiao, Martin Krzywicki, Reta Kutche, Oliver Lee, Soo
Sen Lee, Victor Ling, Carrie Mathewson, Candice McLeavy, Steven
Ness, Pawan Pandon, Anna-Lisa Prabhu, Parvaneh Saeedi, Jacqueline
Schein, Duane Smallus, Michael Smith, Lorraine Spence, Jeff Stott,
Michael Thorne, Miranada Teal, Natasja van den Bosch, Jill Vardy,
George Yang, Scott Zuyderduyn, Marco Marra.

Clone distribution: MGC clone distribution information can be found
through the I.M.A.G.E. Consortium/LNL at: <http://image.lnl.gov>
Series: IRAK Plate: 57 Row: m Column: 21
This clone was selected for full length sequencing because it
passed the following selection criteria: Hexamer frequency ORF
analysis.

FEATURES
source
Location/Qualifiers
1. 1586
/organism="Mus musculus"
/db_xref="taxon:10090"
/map="FVB/N"
/clone="IMAGE:5134400"
/issue_type="Liver, normal, 5 month old male mouse."
/clone_lib="NCI CGAP_L19"
/lab_host="DH10B"
/note="Vector: pCMV-SPORT6"
1. 809
/codon_start=3
/product="unknown (protein for IMAGE:5134400)"
/protein_id="AAH22157.1"
/db_xref="GI:18381058"
/translation="IPDTLTITTELGAVEMVITQLQTSWEDFRICLSIGRLHLHLAH
SPISGVTLIDPRPFLVNGELKLTLDADPVEETPCTSSADCTLEPPARSLPCS
AQGCEGMEKERNLNAVRFPTVLPSPAPSLRLDLSITNATGELAMGVDELTAO
LETAHLPRSGOYLVNTPSSRAVORIRPSAIDODPYRCWPSYHGGCLISFNLALAA
IDVESHQCAFRFVNTQNTWGRKLVFFKIGMNOVDPAGKTTVXAPG"

CDS
BASE COUNT 354 a 455 c 441 g 336 t
ORIGIN

Query Match 2.6%; Score 36; DB 10; Length 1586;
Best Local Similarity 100.0%; Pred. No. 2.3e-09;
Matches 36; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 203 GCCTTTGTGTCACCAACGACGACCTGAGCGTGAGCT 238
|||||
DB 690 GCCTTTGTGTCACCAACGACGACCTGAGCGTGAGCT 725
|||||

RESULT 11
LOCUS AX056461 1886 bp DNA linear PAT 13-JAN-2001
DEFINITION Sequence 105 from Patent WO0073469.

ACCESSION AX056461.1 GI:12229168
 VERSION AX056461.1
 KEYWORDS
 SOURCE
 ORGANISM
 Murinae gen. sp.
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae.
 REFERENCE
 1 (bases 1 to 1886)
 Plozman, G.D., Martinez, R., Whyte, D. and Sudersanam, S.
 TITLE
 Protein kinases
 JOURNAL
 Patent: WO 0073469-A 105 07-DEC-2000;
 Sugen, Inc. (US)
 FEATURES
 source
 1..1886
 Location/Qualifiers
 /organism="Murinae gen. sp."
 /db_xref="taxon:39108"
 BASE COUNT 384 a 559 c 553 g 390 t
 ORIGIN
 Query Match 2.6%; Score 36; DB 6; Length 1886;
 Best Local Similarity 100.0%; Pred. No. 2.3e-09;
 Matches 36; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Oy 203 GCCTTGTGCTACCAACCAAGACCTGACAGGT 238
 Db 981 GCCTTGTGCTACCAACCAAGACCTGACAGGT 1016
 RESULT 12
 AC11157/c
 LOCUS
 DEFINITION
 Homo sapiens chromosome 8 clone RP11-70C11 map 8, LOW-PASS SEQUENCE
 SAMPLING.
 AC11157
 Homo sapiens chromosome 8 clone RP11-70C11 map 8, LOW-PASS SEQUENCE
 AC11157.1 GI:18699922
 HTG: HTGS_PHAS50.
 SOURCE
 Homo sapiens
 ORGANISM
 Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 REFERENCE
 1 (bases 1 to 69492)
 Birren, B., Linton, L., Nusbaum, C. and Lander, E.
 TITLE
 Homo sapiens chromosome 8, clone RP11-70C11
 JOURNAL
 Unpublished
 REFERENCE
 2 (bases 1 to 69492)
 Birren, B., Linton, L., Nusbaum, C., Lander, E., Ali, A., Allen, N.,
 Anderson, S., Barna, N., Bastien, V., Boguslavsky, L., Boukhgalter, B.,
 Brown, A., Camarata, J., Campoliano, A., Chang, J., Chazaro, B.,
 Chopra, Y., Colangelo, M., Collins, S., Collymore, A., Cook, A.,
 Cooke, P., Dearellano, K., Dewar, K., Diaz, J.S., Dodge, S., Faro, S.,
 Ferreira, P., Fitzhugh, W., Gage, D., Galagan, J., Gardyna, S.,
 Ginde, S., Gord, S., Goyette, M., Graham, L., Grand-Pierre, N.,
 Hagos, B., Horton, L., Hulme, W., Iliev, I., Johnson, R., Jones, C.,
 Kamat, A., Karatas, A., Kells, C., Larocque, K., Lamazares, R.,
 Landers, T., Lehoczek, J., Levine, R., Liu, G., Maclean, C.,
 MacDonald, P., Major, J., Margis, N., Matthews, C., McCarthy, M.,
 McKernan, K., Meldrum, J., Meneses, L., Mihova, T.,
 Mlenga, V., Murphy, T., Nayler, J., Nguyen, C., Nicol, R., Notcu, C.,
 Norman, C.H., O'Connor, T., O'Donnell, P., O'Neill, D., Oliver, J.,
 Peterson, K., Phunkhang, P., Pierre, N., Pollara, V., Raymond, C.,
 Retta, R., Rieback, M., Riley, R., Rise, C., Rogov, P., Roman, J.,
 Rosetti, M., Roy, A., Santos, R., Schauer, S., Schnapack, R., Seaman, S.,
 Severy, P., Spencer, B., Stange, R., Stojanovic, N.,
 Strube, N., Sudarmanian, A., Talamas, J., Testafaye, S., Theodore, J.,
 Topham, K., Travers, M., Travis, N., Trigilio, J., Vassiliev, H.,
 Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W. J., Young, G.,
 Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.
 TITLE
 Direct Submissions
 JOURNAL
 Submitted (18-FEB-2002) Whitehead Institute/MIT Center for Genome
 Research, 320 Charles Street, Cambridge, MA 02141, USA
 COMMENT
 All repeats were identified using RepeatMasker:
 Smit, A.F.A. & Green, P. (1996-1997)
 http://ftp.genome.washington.edu/RM/RepeatMasker.html

----- Genome Center
 Center: Whitehead Institute/ MIT Center for Genome Research
 Center code: MIBR
 Web site: http://www-seq.wi.mit.edu
 Contact: sequence_submissions@genome.wi.mit.edu
 Project Information
 Center project name: L25396
 Center clone name: 70_C_11

 * NOTE: This record contains 86 individual
 * sequencing reads that have not been assembled into
 * contigs. Runs of N are used to separate the reads
 * and the order in which they appear is completely
 * arbitrary. Low-pass sequence sampling is useful for
 * identifying clones that may be gene-rich and allows
 * overlap relationships among clones to be deduced.
 * However, it should not be assumed that this clone
 * will be sequenced to completion. In the event that
 * the record is updated, the accession number will
 * be preserved.
 1
 706 805: gap of 100 bp
 806 1512: contig of 707 bp in length
 1513 1612: gap of 100 bp
 1613 2325: contig of 713 bp in length
 2326 2425: gap of 100 bp
 2426 3121: contig of 696 bp in length
 3122 3221: gap of 100 bp
 3222 3930: contig of 709 bp in length
 3931 4030: gap of 100 bp
 4031 4744: contig of 714 bp in length
 4745 4844: gap of 100 bp
 4845 5561: contig of 717 bp in length
 5562 5661: gap of 100 bp
 5662 6390: contig of 729 bp in length
 6391 6490: gap of 100 bp
 6491 7201: contig of 711 bp in length
 7202 7301: gap of 100 bp
 7302 8011: contig of 710 bp in length
 8012 8111: gap of 100 bp
 8112 8829: contig of 718 bp in length
 8830 8929: gap of 100 bp
 8930 9616: contig of 687 bp in length
 9617 9716: gap of 100 bp
 9717 10409: contig of 693 bp in length
 10410 10509: gap of 100 bp
 10510 11220: contig of 711 bp in length
 11221 11320: gap of 100 bp
 11321 12020: contig of 700 bp in length
 12021 12120: gap of 100 bp
 12121 12842: contig of 722 bp in length
 12843 12942: gap of 100 bp
 12943 13641: contig of 699 bp in length
 13642 13741: gap of 100 bp
 13742 14470: contig of 729 bp in length
 14471 14570: gap of 100 bp
 14571 15280: contig of 710 bp in length
 15281 15380: gap of 100 bp
 15381 16097: contig of 717 bp in length
 16098 16197: gap of 100 bp
 16198 16910: contig of 713 bp in length
 16911 17010: gap of 100 bp
 17011 17707: contig of 697 bp in length
 17708 17807: gap of 100 bp
 17808 18509: contig of 702 bp in length
 18510 18609: gap of 100 bp
 18610 19314: contig of 705 bp in length
 19315 19414: gap of 100 bp
 19415 20120: contig of 706 bp in length
 20121 20220: gap of 100 bp
 20221 20925: contig of 705 bp in length
 20926 21025: gap of 100 bp
 21026 21730: contig of 705 bp in length

```

* 21731 21830: gap of 100 bp
* 21831 22540: contig of 710 bp in length
* 22541 22640: gap of 100 bp
* 22641 23362: contig of 722 bp in length
* 23363 23462: gap of 100 bp
* 23463 24186: contig of 724 bp in length
* 24187 24286: gap of 100 bp
* 24287 25003: contig of 717 bp in length
* 25004 25103: gap of 100 bp
* 25104 25812: contig of 709 bp in length
* 25813 25912: gap of 100 bp
* 25913 26629: contig of 717 bp in length
* 26630 26729: gap of 100 bp
* 26730 27427: contig of 698 bp in length
* 27428 27527: gap of 100 bp
* 27528 28249: contig of 722 bp in length
* 28250 28349: gap of 100 bp
* 28350 29065: contig of 716 bp in length
* 29066 29165: gap of 100 bp
* 29166 29868: contig of 703 bp in length
* 29869 29968: gap of 100 bp
* 29969 30686: contig of 718 bp in length
* 30687 30786: gap of 100 bp
* 30787 31496: contig of 710 bp in length
* 31497 31596: gap of 100 bp
* 31597 32358: contig of 762 bp in length
* 32359 32458: gap of 100 bp
* 32459 33171: contig of 713 bp in length
* 33172 33271: gap of 100 bp
* 33272 33982: contig of 711 bp in length
* 33983 34082: gap of 100 bp
* 34083 34769: contig of 687 bp in length
* 34770 34869: gap of 100 bp
* 34870 35573: contig of 704 bp in length
* 35574 35673: gap of 100 bp
* 35674 36382: contig of 709 bp in length
* 36383 36482: gap of 100 bp
* 36483 37180: contig of 698 bp in length
* 37181 37280: gap of 100 bp
* 37281 37983: contig of 703 bp in length
* 37984 38083: gap of 100 bp
* 38084 38799: contig of 716 bp in length
* 38800 38899: gap of 100 bp
* 38900 39611: contig of 712 bp in length
* 39612 39711: gap of 100 bp
* 39712 40424: contig of 713 bp in length
* 40425 40524: gap of 100 bp
* 40525 41242: contig of 718 bp in length
* 41243 41342: gap of 100 bp
* 41343 42052: contig of 710 bp in length
* 42053 42152: gap of 100 bp
* 42153 42857: contig of 705 bp in length
* 42858 42957: gap of 100 bp
* 42958 43670: contig of 713 bp in length
* 43671 43770: gap of 100 bp
* 43771 44477: contig of 707 bp in length
* 44478 44577: gap of 100 bp
* 44578 45283: contig of 706 bp in length
* 45284 45383: gap of 100 bp
* 45384 46067: contig of 684 bp in length
* 46068 46167: gap of 100 bp
* 46168 46877: contig of 710 bp in length
* 46878 46977: gap of 100 bp
* 46978 47660: contig of 683 bp in length
* 47661 47760: gap of 100 bp
* 47761 48476: contig of 716 bp in length
* 48477 48576: gap of 100 bp
* 48577 49290: contig of 714 bp in length
* 49291 49390: gap of 100 bp
* 49391 50111: contig of 721 bp in length
* 50112 50211: gap of 100 bp
* 50212 50913: contig of 702 bp in length
* 50914 51013: gap of 100 bp

```

```

Query Match      2.4%  Score 33; DB 2; Length 69492;
Best Local Similarity 100.0%; Pred.No. 1.4e-07;
Matches 33; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      1359 ACTCTCCCTAAAAA...AAAAAAG 1391
Db      67523 ACTCTCCCTAAAAA...AAAAAAG 67491

RESULT 13
AC091069/c
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM

REFERENCE
AUTHORS
1 (bases 1 to 164796)
Bao,W., Bao,J., Bao,Q., Bian,X., Cao,T., Chen,C., Chen,J., Ding,H.,
He,L., Hu,S., Huang,F., Jin,Y., Kang,N., Li,C., Li,C., Li,F.,
Li,G., Li,J., Li,L., Li,S., Li,T., Liu,Y., Liu,N., Liu,B., Liu,Y.,
Li,W., Li,W., Li,Y., Luo,J., Niu,Y., Qi,Q., Qi,X., Song,L.,
Song,S., Sun,M., Sun,W., Sun,Y., Tan,X., Tao,R., Wang,H., Wang,J.,
Wang,J., Wang,J., Wang,L., Wang,L., Wang,R., Wang,X., Wang,X.,
Wang,Y., Wu,D., Wu,Q., Xie,F., Xuan,Z., Xue,Y., Yan,C., Yang,X.,
Yu,B., Zeng,Y., Zhang,G., Zhang,H., Zhang,H., Zhang,L., Zhang,M.,
Zhang,X., Zhang,X., Zhang,Y., Zhang,Y., Zhang,Z., Zhu,B., Zhu,N.,
Yu,J. and Yang,H.
Chromosome 3p genomic sequence
Unpublished
2 (bases 1 to 164796)
Bao,W., Bao,J., Bao,Q., Bian,X., Cao,T., Chen,C., Chen,J., Ding,H.,
Dong,W., Fan,H., Feng,X., Gong,J., Guan,Q., Gu,X., Guo,D., Guo,Z.,
He,L., Hu,S., Huang,F., Jin,Y., Kang,N., Li,C., Li,C., Li,F.,
Li,G., Li,J., Li,L., Li,S., Li,T., Liu,Y., Liu,N., Liu,B., Liu,Y.,
Li,W., Li,W., Li,Y., Luo,J., Niu,Y., Qi,Q., Qi,X., Song,L.,
Song,S., Sun,M., Sun,W., Sun,Y., Tan,X., Tao,R., Wang,H., Wang,J.,
Wang,J., Wang,J., Wang,L., Wang,L., Wang,R., Wang,X., Wang,X.,
Wang,Y., Wu,D., Wu,Q., Xie,F., Xuan,Z., Xue,Y., Yan,C., Yang,X.,
Yu,B., Zeng,Y., Zhang,G., Zhang,H., Zhang,H., Zhang,L., Zhang,M.,
Zhang,X., Zhang,X., Zhang,Y., Zhang,Y., Zhang,Z., Zhu,B., Zhu,N.,
Yu,J. and Yang,H.
Chromosome 3p genomic sequence
Unpublished
TITLE
JOURNAL
AUTHORS
COMMENT
Genome Center
Center:Beijing
Center code:Beijing
Website:http://hgsc.igbp.ac.cn
http://www.genomics.org.cn
Contact:hgsc@igbp.ac.cn
Project Information
Center project name:1# project
Center clone name: Rpl1-16113
Summary Statistics

```

```
misc_feature /clone lib=RPCT-11 Human Male BAC
1. .2517
/notes="assembly_fragment"
2618. .5699
/notes="assembly_fragment
clone end:SP6
vector side:right"
5800.7367
/notes="assembly_fragment"
9468. .17246
/notes="assembly_fragment
clone end:T7
vector side:right"
17347. .28325
/notes="assembly_fragment"
28426. .41647
/notes="assembly_fragment"
41748. .55693
/notes="assembly_fragment"
55794. .79643
/notes="assembly_fragment"
79744. .120543
/notes="assembly_fragment"
120644. .165201
/notes="assembly_fragment"
```

BASE COUNT 53047 a 31498 c 31586 g 48170 t 900 others
 ORIGIN

Query Match 2.3%; Score 32; DB 2; Length 165201;
 Best Local Similarity 100.0%; Pred. No. 5.7e-07;
 Matches 32; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1358 AACTCTCCCTAATAAAAAAAAAAAAAAAAAA 1389
 |||
 DB 33306 AACTCTCCCTAATAAAAAAAAAAAAAAAAAA 33275

RESULT 15
 HSA420562 1416 bp mRNA linear PRI 23-NOV-2001
 LOCUS Homo sapiens mRNA full length insert cDNA clone EUROIMAGE 1525273.
 DEFINITION
 ACCESSION AJ420562
 VERSION AJ420562.1 GI:17066426
 KEYWORDS
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE 1
 Auffray C., Ansoerge W., Ballabio A., Estivill X., Gibson K.,
 Leitch H., Poustka A. and Lundberg U.
 TITLE The European IMAGE consortium for integrated Molecular analysis of
 human gene transcripts
 JOURNAL Unpublished
 REFERENCE 2 (bases 1 to 1416)
 AUTHORS Persson A.
 TITLE Direct Submission
 JOURNAL Submitted (02-OCT-2001) Persson A., Center for Molecular
 Biotechnology, KTH, SCFAB, Institute of Biotechnology, Roslagsevagen
 30B, 106 91 Stockholm, SWEDEN

COMMENT This clone is available royalty-free through IMAGE Consortium
 distributors. IMPORTANT: This sequence represents the full insert
 of this IMAGE cDNA clone. No attempt has been made to verify
 whether this corresponds to the full-length of the original mRNA
 from which it was derived.

FEATURES
 source
 1..1416
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /chromosome="6"
 /clone="EUROIMAGE 1525273"
 /clone_1lb="Soares_NFL_T_GBC_S1"
 polyA_signal 1357..1362
 BASE COUNT 484 a 237 c 276 g 419 t
 ORIGIN

Query Match 2.3%; Score 30; DB 9; Length 1416;
 Best Local Similarity 100.0%; Pred. No. 7.8e-06;
 Matches 30; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1361 TCTTCCTAATAAAAAAAAAAAAAAAAAA 1390
 |||
 DB 1375 TCTTCCTAATAAAAAAAAAAAAAAAAAA 1404

Search completed: June 29, 2003, 06:12:36
 Job time : 2538 secs

C 229	26	1.9	360	22	AA10196	Human breast cancer	302	26	1.9	981	20	AAK87394	Mouse STR1F1 (Tan
C 230	26	1.9	360	22	AA11976	Human breast cancer	303	26	1.9	1040	24	ABL56860	Mouse X-type secre
C 231	26	1.9	361	23	ABV12942	Human prostate exp	304	26	1.9	1040	24	ABA96630	Mouse X-type secre
C 232	26	1.9	365	23	ABV47637	Human prostate exp	305	26	1.9	1062	22	AAH31412	Human colon cancer
C 233	26	1.9	366	22	AA187465	Human polynucleoti	306	26	1.9	1242	22	AAD02808	HNA040 CDNA/clon
C 234	26	1.9	372	24	ABK64383	Human benign prost	307	26	1.9	1263	21	AAFI5596	Human prostate can
C 235	26	1.9	374	22	AA181153	Human polynucleoti	308	26	1.9	1274	24	ABD38087	Corn chlorophyllas
C 236	26	1.9	382	23	ABV46589	Human prostate exp	309	26	1.9	1318	24	ABD35068	Marine CDNA isolat
C 237	26	1.9	387	22	AAH71915	Human cervical can	310	26	1.9	1327	20	AAK30162	Human secreted pro
C 238	26	1.9	391	22	AA183015	Human polynucleoti	311	26	1.9	1391	21	AACT7672	Human cancer/assoc
C 239	26	1.9	396	23	ABV49020	Human prostate exp	312	26	1.9	1395	24	ABO51152	Oligonucleotide fo
C 240	26	1.9	397	22	AA180153	Human polynucleoti	313	26	1.9	1395	24	ABO51152	Oligonucleotide fo
C 241	26	1.9	401	22	AA183850	Human polynucleoti	314	26	1.9	1431	22	AAD20262	Rat integrin modu
C 242	26	1.9	402	22	AA187524	Human polynucleoti	315	26	1.9	1509	22	AA158831	Human secreted pro
C 243	26	1.9	402	22	AA110391	Human breast cancer	316	26	1.9	1562	24	ABK69868	Human secreted pro
C 244	26	1.9	403	22	AA185357	Human polynucleoti	317	26	1.9	1562	24	ABK69868	Human secreted pro
C 245	26	1.9	403	23	ABV61005	Human prostate exp	318	26	1.9	1569	21	AAZ98057	Human secreted pro
C 246	26	1.9	405	22	AA119599	Human breast cancer	319	26	1.9	1569	22	AAD11670	Human secreted pro
C 247	26	1.9	405	23	ABV34066	Human prostate exp	320	26	1.9	1569	24	ABK69766	Human secreted pro
C 248	26	1.9	405	23	ABV42931	Human prostate exp	321	26	1.9	1728	23	AAK87014	DNA encoding novel
C 249	26	1.9	413	22	AAK43374	DNA encoding novel	322	26	1.9	1730	21	AAA26383	Human secreted pro
C 250	26	1.9	413	22	AA189438	Human polynucleoti	323	26	1.9	1758	17	AAAT32325	Soluble starch syn
C 251	26	1.9	416	23	ABV47951	Human prostate exp	324	26	1.9	1797	21	AAK08375	Rat 4F2 heavy chai
C 252	26	1.9	423	22	AA182334	Human polynucleoti	325	26	1.9	1852	22	AAH18681	Human CDNA sequenc
C 253	26	1.9	427	22	AAD05269	Human secreted pro	326	26	1.9	1960	24	ABN87244	Lolium perenne C4H
C 254	26	1.9	433	22	AA1823127	Human breast cancer	327	26	1.9	2093	17	AAAT17864	Hepatitis A virus
C 255	26	1.9	433	22	AA187368	Human polynucleoti	328	26	1.9	2123	24	ABL59578	Rat OST26 gene fra
C 256	26	1.9	437	22	AA118755	Human breast cancer	329	26	1.9	2123	24	ABL59578	Rat OST29 gene fra
C 257	26	1.9	440	22	AAH70715	Human cervical can	330	26	1.9	2158	24	ABL50337	Human cancer cell
C 258	26	1.9	441	22	AA119793	Human breast cancer	331	26	1.9	2208	21	AAK59775	Human secreted pro
C 259	26	1.9	442	22	AA183061	Human polynucleoti	332	26	1.9	2268	21	AAK66441	Human secreted pro
C 260	26	1.9	448	22	AA186117	Human polynucleoti	333	26	1.9	2382	22	AA197952	Human neuroblastom
C 261	26	1.9	448	23	ABV49465	Human prostate exp	334	26	1.9	2362	22	AA198105	Human neuroblastom
C 262	26	1.9	448	24	AB187531	Human ovarian can	335	26	1.9	2528	21	AAFI6147	Human prostate can
C 263	26	1.9	453	22	AA119861	Human breast cancer	336	26	1.9	2535	22	AAK41164	CDNA encoding nove
C 264	26	1.9	471	24	ABN87880	Human ovary specif	337	26	1.9	2635	22	ABK43877	DNA encoding novel
C 265	26	1.9	473	22	AA190495	Human polynucleoti	338	26	1.9	2660	23	AAH43560	Human DHOR-1 CDNA
C 266	26	1.9	482	23	ABV49211	Human prostate exp	339	26	1.9	2745	23	ABK43576	DNA encoding novel
C 267	26	1.9	504	22	AAK57962	Human prostate exp	340	26	1.9	3001	24	ABK34025	Human DNA for strag
C 268	26	1.9	511	21	AAZ97335	Human immune/haema	341	26	1.9	3149	24	ABK84687	Human CDNA differe
C 269	26	1.9	531	23	ABV46374	Human prostate can	342	26	1.9	3260	22	AAD07596	Human secreted pro
C 270	26	1.9	534	23	ABV47036	Human prostate exp	343	26	1.9	3289	21	AAK26281	Human secreted pro
C 271	26	1.9	540	23	ABV60963	Human prostate exp	344	26	1.9	3304	22	ABAI18014	Human nervous syst
C 272	26	1.9	549	20	AAK87732	Beauveria bassiana	345	26	1.9	3338	21	AAZ94119	Haematopoietic ste
C 273	26	1.9	578	23	ABV54931	Human prostate exp	346	26	1.9	3408	21	AAZ94096	Haematopoietic ste
C 274	26	1.9	580	22	AAH70713	Human cervical can	347	26	1.9	3536	22	AAK06381	Human ATP binding
C 275	26	1.9	594	24	AA137818	Wheat KCP-like pro	348	26	1.9	3591	20	AAK60932	Murine cell death
C 276	26	1.9	595	22	AAK56873	Human CDNA encodin	349	26	1.9	3712	20	AAK03040	Human IL-1ra BAC c
C 277	26	1.9	600	24	ABO57191	Human colon cancer	350	26	1.9	4116	20	AAK02983	Human IL-1ra BAC c
C 278	26	1.9	619	24	ABO22348	Oligonucleotide fo	351	26	1.9	5640	24	AA169391	Human EDG-8 recept
C 279	26	1.9	619	24	ABO22349	Oligonucleotide fo	352	26	1.9	5647	24	ABL70355	Chemically treated
C 280	26	1.9	631	24	ABO25138	Oligonucleotide fo	353	26	1.9	5647	24	ABL33566	Human immune syste
C 281	26	1.9	631	24	ABO25139	Oligonucleotide fo	354	26	1.9	5647	24	AAK561320	Human immune syste
C 282	26	1.9	632	24	ABO60764	Human colon cancer	355	26	1.9	6086	22	AAK46667	Human gene regulat
C 283	26	1.9	639	23	ABV19696	Human prostate exp	356	26	1.9	6165	24	ABN97371	Tumour suppressor
C 284	26	1.9	653	24	ABO57371	Human colon cancer	357	26	1.9	6352	24	ABL32904	Gene #1869 used to
C 285	26	1.9	655	24	AB187011	Human ovarian can	358	26	1.9	6484	24	ABN80138	Human immune syste
C 286	26	1.9	700	22	AAH92102	Human inflammatory	359	26	1.9	6553	22	AAK40055	Human chemically m
C 287	26	1.9	711	20	AAK80828	Human secreted pro	360	26	1.9	6553	22	AAK91472	Genomic sequence #
C 288	26	1.9	711	21	AAK52476	Human secreted pro	361	26	1.9	6668	24	ABL33697	Human digestive sy
C 289	26	1.9	755	24	ABO89771	Human prostate exp	362	26	1.9	6686	24	ABL33697	Human immune syste
C 290	26	1.9	784	24	AAK62079	Porcine muscular s	363	26	1.9	6854	22	AA137419	Human immune syste
C 291	26	1.9	785	22	AAK95240	Human neuroblastom	364	26	1.9	7921	24	ABL33970	Human muscle skele
C 292	26	1.9	785	22	AA198135	Human neuroblastom	365	26	1.9	8912	16	AAAT0040	Human immune syste
C 293	26	1.9	798	22	AA196495	Human neuroblastom	366	26	1.9	8912	16	AAK55280	Hepatitis GB virus
C 294	26	1.9	875	24	ABQ24158	Oligonucleotide fo	367	26	1.9	9139	22	AAK23485	Hepatitis GB virus
C 295	26	1.9	875	24	ABQ24159	Oligonucleotide fo	368	26	1.9	9143	18	AAAT00130	Hepatitis GB virus
C 296	26	1.9	889	22	AA197724	Human neuroblastom	369	26	1.9	9143	16	AAAT59785	Hepatitis GB virus
C 297	26	1.9	899	20	ABAI19428	Nicotiana tabacum	370	26	1.9	9143	21	AAK55376	Hepatitis GB virus
C 298	26	1.9	920	22	ABAI18016	Human nervous syst	371	26	1.9	9143	21	AAK55376	Hepatitis GB virus
C 299	26	1.9	927	24	ABT03001	Human breast spect	372	26	1.9	9399	22	AAK23484	GBV-B virus genome
C 300	26	1.9	944	24	ABK84455	Human CDNA differe	373	26	1.9	10891	22	ABL32464	Human immune syste
C 301	26	1.9	944	24	ABN98582	Arabidopsis thalia	374	26	1.9	11744	24	AAK83160	Human immune/haema

C 521	25	1.8	377	23	ABV10381	Human prostate exp
C 522	25	1.8	379	22	AA183173	Human polynucleoti
C 523	25	1.8	381	22	AAK61735	Human immune/haema
C 524	25	1.8	381	23	ABV58662	Human prostate exp
C 525	25	1.8	381	24	AB085304	Arabidopsis thalia
C 526	25	1.8	382	22	AA180890	Human polynucleoti
C 527	25	1.8	382	22	AA182890	Human polynucleoti
C 528	25	1.8	382	22	AA183090	Human polynucleoti
C 529	25	1.8	382	22	AA184706	Human polynucleoti
C 530	25	1.8	382	22	AA185598	Human polynucleoti
C 531	25	1.8	382	22	AA187806	Human polynucleoti
C 532	25	1.8	382	23	ABV49116	Human prostate exp
C 533	25	1.8	383	22	AA188525	Human polynucleoti
C 534	25	1.8	384	21	AA189598	Human secreted pro
C 535	25	1.8	384	22	AA183079	Human polynucleoti
C 536	25	1.8	384	22	AA191477	Human polynucleoti
C 537	25	1.8	385	22	AA187561	Human polynucleoti
C 538	25	1.8	386	22	AA189019	Human polynucleoti
C 539	25	1.8	386	22	AA192181	Human polynucleoti
C 540	25	1.8	388	23	ABV00188	Human prostate exp
C 541	25	1.8	389	22	AA180882	Human breast cance
C 542	25	1.8	389	23	ABV61827	Human prostate exp
C 543	25	1.8	390	22	AA184812	Human polynucleoti
C 544	25	1.8	390	22	AA188511	Human polynucleoti
C 545	25	1.8	391	22	AA186854	Human polynucleoti
C 546	25	1.8	394	22	AA196958	Human Mammary Gland
C 547	25	1.8	395	22	AA181559	Human polynucleoti
C 548	25	1.8	396	22	AA188643	Human polynucleoti
C 549	25	1.8	398	21	AA1844860	Human secreted exp
C 550	25	1.8	398	22	AA188703	Human polynucleoti
C 551	25	1.8	399	22	AA183476	Human polynucleoti
C 552	25	1.8	400	22	AA185595	Human polynucleoti
C 553	25	1.8	400	22	AA187360	Human polynucleoti
C 554	25	1.8	400	22	AA187455	Human polynucleoti
C 555	25	1.8	401	22	AAK96707	Human neurotulin g
C 556	25	1.8	401	22	AAK98200	Human neurotulin g
C 557	25	1.8	401	22	AA183877	Human polynucleoti
C 558	25	1.8	403	22	AA182884	Human polynucleoti
C 559	25	1.8	404	22	AA183869	Human polynucleoti
C 560	25	1.8	404	23	ABV14862	Human prostate exp
C 561	25	1.8	404	23	ABV49258	Human prostate exp
C 562	25	1.8	405	22	AAK96652	Human prostate exp
C 563	25	1.8	405	22	AAK98145	Human neurotulin g
C 564	25	1.8	405	22	AA184779	Human polynucleoti
C 565	25	1.8	405	23	ABV341801	Human prostate exp
C 566	25	1.8	405	23	ABV34041	Human prostate exp
C 567	25	1.8	406	19	AAV05722	Nucleotide sequenc
C 568	25	1.8	406	22	AA181228	Human polynucleoti
C 569	25	1.8	406	22	AA185509	Human polynucleoti
C 570	25	1.8	406	22	AA192822	Human polynucleoti
C 571	25	1.8	407	23	ABV46364	Human prostate exp
C 572	25	1.8	407	23	ABV61484	Human prostate exp
C 573	25	1.8	408	22	AAH10289	Human CDNA clone (
C 574	25	1.8	408	23	ABV49309	Human prostate exp
C 575	25	1.8	409	22	AA180756	Human polynucleoti
C 576	25	1.8	409	22	AA184857	Human polynucleoti
C 577	25	1.8	410	23	ABV39040	Human prostate exp
C 578	25	1.8	410	23	ABV44826	Human prostate exp
C 579	25	1.8	410	23	ABV48387	Human prostate exp
C 580	25	1.8	411	22	AA184864	Human polynucleoti
C 581	25	1.8	411	22	AA187361	Human polynucleoti
C 582	25	1.8	411	22	AA187378	Human polynucleoti
C 583	25	1.8	412	22	AA184848	Human polynucleoti
C 584	25	1.8	412	23	ABV43247	Human prostate exp
C 585	25	1.8	413	22	AA182940	Human polynucleoti
C 586	25	1.8	413	22	AA184095	Human polynucleoti
C 587	25	1.8	413	22	AA185119	Human polynucleoti
C 588	25	1.8	413	23	ABV14974	Human prostate exp
C 589	25	1.8	414	21	AA181730	Human polynucleoti
C 590	25	1.8	415	22	ABV60254	Human prostate exp
C 591	25	1.8	416	22	AA191796	Human polynucleoti
C 592	25	1.8	416	22	AA191796	Human polynucleoti
C 593	25	1.8	418	21	AA1938309	Human colon cancer
C 594	25	1.8	418	23	ABV14865	Human prostate exp
C 595	25	1.8	422	22	AA187575	Human polynucleoti
C 596	25	1.8	422	23	ABV43774	Human prostate exp
C 597	25	1.8	424	22	AA187099	Human polynucleoti
C 598	25	1.8	425	21	AA187099	Human polynucleoti
C 599	25	1.8	425	21	AA187099	Human polynucleoti
C 600	25	1.8	430	21	AA187099	Human polynucleoti
C 601	25	1.8	430	21	AA187099	Human polynucleoti
C 602	25	1.8	433	22	AA183988	Human polynucleoti
C 603	25	1.8	433	22	AA183988	Human polynucleoti
C 604	25	1.8	433	22	AA183988	Human polynucleoti
C 605	25	1.8	438	22	AA182079	Human prostate exp
C 606	25	1.8	438	22	AA182079	Human prostate exp
C 607	25	1.8	438	22	AA182079	Human prostate exp
C 608	25	1.8	440	22	AA183772	Human polynucleoti
C 609	25	1.8	440	22	AA183772	Human polynucleoti
C 610	25	1.8	442	22	AA191014	Human polynucleoti
C 611	25	1.8	442	22	AA180488	Human polynucleoti
C 612	25	1.8	442	22	AA180488	Human polynucleoti
C 613	25	1.8	442	22	AA180488	Human polynucleoti
C 614	25	1.8	447	23	ABV13449	Human prostate exp
C 615	25	1.8	448	22	AA191812	Human polynucleoti
C 616	25	1.8	448	23	ABV19167	Human prostate exp
C 617	25	1.8	449	22	AA110213	Human breast cance
C 618	25	1.8	449	22	AA183053	Human prostate exp
C 619	25	1.8	450	22	AA183053	Human prostate exp
C 620	25	1.8	451	23	ABV38949	Human polynucleoti
C 621	25	1.8	451	23	ABV38949	Human prostate exp
C 622	25	1.8	451	24	ABV38949	Human prostate exp
C 623	25	1.8	451	24	ABV38949	Human prostate exp
C 624	25	1.8	454	22	AA185101	Human CDNA for an
C 625	25	1.8	454	22	AA185101	Human polynucleoti
C 626	25	1.8	454	23	ABV14338	Human prostate exp
C 627	25	1.8	454	23	ABV39057	Human prostate exp
C 628	25	1.8	461	22	AA110372	Human breast cance
C 629	25	1.8	461	23	ABV39039	Human prostate exp
C 630	25	1.8	461	23	ABV44778	Human prostate exp
C 631	25	1.8	461	23	ABV44825	Human prostate exp
C 632	25	1.8	463	22	AA183750	Human polynucleoti
C 633	25	1.8	463	23	ABV18369	Human prostate exp
C 634	25	1.8	464	22	AA183736	Human polynucleoti
C 635	25	1.8	464	23	ABV34269	Human prostate exp
C 636	25	1.8	464	23	ABV41330	Human prostate exp
C 637	25	1.8	466	22	AA182085	Human polynucleoti
C 638	25	1.8	467	23	ABV45888	Human prostate exp
C 639	25	1.8	469	24	ABK45850	CDNA encoding colo
C 640	25	1.8	470	22	AA125852	Human breast cance
C 641	25	1.8	470	22	AA193681	Human polynucleoti
C 642	25	1.8	470	23	ABV14765	Human prostate exp
C 643	25	1.8	470	23	ABV49367	Human prostate exp
C 644	25	1.8	470	23	ABV61721	Human prostate exp
C 645	25	1.8	472	23	ABV46233	Human prostate exp
C 646	25	1.8	475	23	ABV30533	Human prostate exp
C 647	25	1.8	475	23	ABV39505	Human prostate exp
C 648	25	1.8	477	22	AA181157	Human polynucleoti
C 649	25	1.8	477	21	AA181157	Human breast land o
C 650	25	1.8	478	23	ABV35326	Human prostate exp
C 651	25	1.8	478	23	ABV44157	Human prostate exp
C 652	25	1.8	481	23	ABV44745	Human prostate exp
C 653	25	1.8	484	23	ABV05169	Human prostate exp
C 654	25	1.8	484	23	ABV59185	Human prostate exp
C 655	25	1.8	485	22	AA116432	Human breast cance
C 656	25	1.8	486	23	ABV35423	Human prostate exp
C 657	25	1.8	487	23	ABV46451	Human prostate exp
C 658	25	1.8	489	22	AA119902	Human breast cance
C 659	25	1.8	493	22	AA119017	Human breast cance
C 660	25	1.8	495	23	ABV18011	Human prostate exp
C 661	25	1.8	496	23	ABV48704	Human prostate exp
C 662	25	1.8	499	22	AAH12544	Human CDNA clone (
C 663	25	1.8	499	23	ABV46838	Human prostate exp
C 664	25	1.8	501	22	AA121795	Human breast cance
C 665	25	1.8	501	22	AA193675	Human polynucleoti
C 666	25	1.8	502	22	AA114084	Human nervous syst

c 667	25	1.8	506	21	AA280662	Human colon cancer	c 740	25	1.8	650	22	AA101722	Human reproductive
c 668	25	1.8	510	23	ABV48680	Human prostate exp	c 741	25	1.8	653	23	ABV05454	Human prostate exp
c 669	25	1.8	510	24	ABN87806	Human ovary specif	c 742	25	1.8	655	20	AA217693	Human gene express
c 670	25	1.8	510	24	ABL83785	Human ovarian canc	c 743	25	1.8	655	21	AA402687	Human colon cancer
c 671	25	1.8	515	19	AAV23109	CDNA encoding the	c 744	25	1.8	657	20	AAV37547	Human secreted pro
c 672	25	1.8	516	22	AAI25122	Human breast cancer	c 745	25	1.8	660	23	ABV14955	Human prostate exp
c 673	25	1.8	516	23	ABV47663	Human prostate exp	c 746	25	1.8	661	24	ABO59445	Human colon cancer
c 674	25	1.8	523	18	AAI91165	Human cacthelexia-ss	c 747	25	1.8	689	22	AA196694	Human neuroblastom
c 675	25	1.8	525	22	ABA13577	Human nervous syst	c 748	25	1.8	690	22	AA125275	Human breast cancer
c 676	25	1.8	530	22	AAH10156	Human CDNA clone (c 749	25	1.8	695	22	AA196733	Human neuroblastom
c 677	25	1.8	533	22	AAH34401	Human colon cancer	c 750	25	1.8	700	22	AAH92947	Human inflammatory
c 678	25	1.8	533	22	AAH10606	Human CDNA clone (c 751	25	1.8	701	22	AAH13166	Human nervous syst
c 679	25	1.8	535	24	AAH04093	Human ovary specif	c 752	25	1.8	703	22	AAI42577	Partial sequence o
c 680	25	1.8	537	23	ABV60414	Human prostate exp	c 753	25	1.8	705	24	AB198562	Human polynucleoti
c 681	25	1.8	539	23	ABV57689	Human prostate exp	c 754	25	1.8	707	24	AB198562	Human neuroblastom
c 682	25	1.8	540	22	ABF81801	Human secreted pro	c 755	25	1.8	710	22	AAI97812	Human neuroblastom
c 683	25	1.8	551	24	ABO58776	Human colon cancer	c 756	25	1.8	710	22	AAI97860	Human neuroblastom
c 684	25	1.8	552	24	ABO60515	Human colon cancer	c 757	25	1.8	710	22	AAI97870	Human neuroblastom
c 685	25	1.8	552	24	AB165284	Lung cancer relate	c 758	25	1.8	711	22	AAI97858	Oligonucleotide fo
c 686	25	1.8	554	23	ABV44736	Human prostate exp	c 759	25	1.8	713	24	ABO28962	Oligonucleotide fo
c 687	25	1.8	557	24	ABO58303	Human colon cancer	c 760	25	1.8	713	24	ABO28963	Human ovary specif
c 688	25	1.8	558	23	ABV35701	Human prostate exp	c 761	25	1.8	716	24	ABO57607	Human colon cancer
c 689	25	1.8	558	23	ABV44506	Human prostate exp	c 762	25	1.8	716	24	AA596302	Arabidopsis CDNA e
c 690	25	1.8	560	23	ABV47703	Human prostate exp	c 763	25	1.8	723	24	AA596330	Human neuroblastom
c 691	25	1.8	563	22	AAI72800	Secreted protein g	c 764	25	1.8	723	24	AA596330	Human neuroblastom
c 692	25	1.8	566	21	AA280382	Human colon cancer	c 765	25	1.8	725	22	AA195619	Human neuroblastom
c 693	25	1.8	568	24	ABK54673	Human colon cancer	c 766	25	1.8	731	22	AA195652	Human neuroblastom
c 694	25	1.8	571	21	AACT7871	Human cancer assoc	c 767	25	1.8	732	22	AA197146	Human neuroblastom
c 695	25	1.8	572	20	AAK63372	Soybean threonine	c 768	25	1.8	736	22	AAH04651	Human CDNA clone (
c 696	25	1.8	573	22	AAK63372	Human immune/haema	c 769	25	1.8	739	22	AAI96968	Human neuroblastom
c 697	25	1.8	576	21	AAI16432	Human ISF CDNA clo	c 770	25	1.8	743	22	AAI97514	Human secreted pro
c 698	25	1.8	580	22	AAK92918	Oligonucleotide fo	c 771	25	1.8	745	22	AAH34348	Human colon cancer
c 699	25	1.8	583	24	ABO52812	Human prostate exp	c 772	25	1.8	747	22	AAI97127	Human neuroblastom
c 700	25	1.8	587	22	ABO62563	Human foecal liver	c 773	25	1.8	747	22	AAI97540	Human neuroblastom
c 701	25	1.8	587	22	ABA29891	Probe #8357 for ge	c 774	25	1.8	751	22	AAI97533	Human neuroblastom
c 702	25	1.8	587	22	AAK10919	Human brain expres	c 775	25	1.8	757	22	AAI24366	Human breast cancer
c 703	25	1.8	587	22	AAK36776	Human bone marrow	c 776	25	1.8	759	22	AAI97528	Human neuroblastom
c 704	25	1.8	587	22	AAI17627	Probe #17560 for ge	c 777	25	1.8	766	22	AAI96361	Oligonucleotide fo
c 705	25	1.8	587	22	AAI42550	Human genome-deriv	c 778	25	1.8	769	22	ABO18554	Oligonucleotide fo
c 706	25	1.8	587	22	ABO510780	Human CDNA for nov	c 779	25	1.8	769	24	ABO18555	Human neuroblastom
c 707	25	1.8	592	24	ABK34545	Human colon cancer	c 780	25	1.8	771	22	AAI97078	Human neuroblastom
c 708	25	1.8	593	21	AA280444	Human colon cancer	c 781	25	1.8	774	24	ABO52344	Oligonucleotide fo
c 709	25	1.8	601	24	ABO59526	Human colon cancer	c 782	25	1.8	774	24	ABO52345	Oligonucleotide fo
c 710	25	1.8	602	22	ABO80820	Human ovarian and	c 783	25	1.8	774	24	ABO52345	Oligonucleotide fo
c 711	25	1.8	602	22	AA106681	Human reproductive	c 784	25	1.8	777	22	AAH83115	Human neuroblastom
c 712	25	1.8	604	20	AAV79062	Human secreted pro	c 785	25	1.8	777	22	AAI96237	Human neuroblastom
c 713	25	1.8	606	23	ABV03945	Human prostate exp	c 786	25	1.8	779	16	AAO79737	Flower style-speci
c 714	25	1.8	606	23	ABV44739	Human prostate exp	c 787	25	1.8	779	18	AAI64554	Tomato S-ribonucle
c 715	25	1.8	606	23	ABV44739	Human prostate exp	c 788	25	1.8	783	22	AAI24838	Human breast cancer
c 716	25	1.8	606	23	ABV44739	Human prostate exp	c 789	25	1.8	785	22	AAI95872	Human neuroblastom
c 717	25	1.8	608	13	AAO31796	Human prostate exp	c 790	25	1.8	788	24	ABO27550	Oligonucleotide fo
c 718	25	1.8	610	22	AAI19953	Human breast cance	c 791	25	1.8	788	24	ABO27551	Oligonucleotide fo
c 719	25	1.8	611	21	AAAI6352	Human colon cancer	c 792	25	1.8	791	22	AAI96087	Human neuroblastom
c 720	25	1.8	612	24	ABO60507	Human colon cancer	c 793	25	1.8	792	22	AAI95856	Human neuroblastom
c 721	25	1.8	612	24	AB187317	Human ovarian canc	c 794	25	1.8	794	23	AA578134	DNA encoding novel
c 722	25	1.8	615	22	AAK88889	Human digestive sy	c 795	25	1.8	795	22	AAH18424	Human CDNA sequenc
c 723	25	1.8	615	22	AAI57610	Human colorectal c	c 796	25	1.8	795	22	AB190186	Human polynucleoti
c 724	25	1.8	618	24	ABN73814	Bovine embryonic g	c 797	25	1.8	797	22	AAI96443	Human neuroblastom
c 725	25	1.8	621	22	AA529124	CDNA encoding for	c 798	25	1.8	802	22	AAI94794	Human neuroblastom
c 726	25	1.8	625	24	ABO57599	Human colon cancer	c 799	25	1.8	805	22	AAI97441	Human neuroblastom
c 727	25	1.8	628	23	ABV23066	Human prostate exp	c 800	25	1.8	809	22	AAK82491	Human immune/haema
c 728	25	1.8	628	23	ABV23066	Human prostate exp	c 801	25	1.8	820	22	AAH04384	Human CDNA clone (
c 729	25	1.8	628	23	ABV28902	Human prostate exp	c 802	25	1.8	828	22	AA196002	Human neuroblastom
c 730	25	1.8	628	23	ABV28902	Human prostate exp	c 803	25	1.8	828	24	ABO14816	Oligonucleotide fo
c 731	25	1.8	631	24	ABV61919	Human prostate exp	c 804	25	1.8	828	24	ABO14817	Oligonucleotide fo
c 732	25	1.8	632	24	ABK62327	Rat sequence diffe	c 805	25	1.8	828	24	ABO46914	Oligonucleotide fo
c 733	25	1.8	635	21	AAAC98042	Human colon cancer	c 806	25	1.8	829	24	AAH04291	Human CDNA clone (
c 734	25	1.8	637	22	AA525985	Human CDNA encodin	c 807	25	1.8	829	22	AAZ25123	Rice trypanophan sy
c 735	25	1.8	640	22	AA100084	Human reproductive	c 808	25	1.8	831	20	AAH64773	Human secreted pro
c 736	25	1.8	643	22	AA556406	Human CDNA for an	c 809	25	1.8	831	22	AAH64772	Human secreted pro
c 737	25	1.8	644	22	ABV13869	Human prostate exp	c 810	25	1.8	832	22	AAI97461	Human neuroblastom
c 738	25	1.8	645	23	ABV14956	Human prostate exp	c 811	25	1.8	833	22	AAI97461	Human neuroblastom
c 739	25	1.8	646	20	AAV79022	Human secreted pro	c 812	25	1.8	842	22	AAI94769	Human neuroblastom

Human ovarian anti
Rat secreted facto
Human colon cancer
C. purpureus delta
cDNA encoding nove
Human fecal brain
Human secreted pro
Human colon cancer
Human secreted pro
Human polynucleoti
Human lung-specific
Human colon cancer
Human neuroblastom
Human neuroblastom
Human nervous syst
Human CDNA for nov
F-box protein FMD1
Human CDNA sequenc
Human prostate exp
Human nervous syst
Human CDNA encodi
Human polynucleoti
Human nervous syst
Human cancer assoc
Human ovarian anti
Human neuroblastom
Human neuroblastom
DNA encoding novel
Human polynucleoti
Hepelocellular car
Human neuroblastom
Human brain CNG-1
Human cancer assoc
Gp85-97 clone 18 c
Human CDNA encodi
Human reproductive
Human secreted pro
Wheat viviparus 1
AAH33227
Human colon cancer
Human pancreatic c

ALIGNMENTS

1000 25 1.8 2342 21 AAC99083

RESULT 1

AA199052 standard; DNA; 9839 BP.

AA199052;

07-JAN-2002 (first entry)

Human excretory related polynucleotide SEQ ID NO 816.

Human; nootropic; neuroprotective; cytostatic; dermatological; vituicide;
immunosuppressive; antiinflammatory; anti-HIV; antibacterial; vulnery;
antiparkinsonian; antischizoid; antianemic; antitubercular; cancer;
antiparasitic; hepatoprotective; cerebroprotective; antiinflammatory;
antiallergic; antidiabetic; antitumor; anticonvulsant; antifungal;
antiparasitic; cardiac; immune disorder; cardiovascular disorder;
neurological disease; infection; nephrotoxic; gene therapy; vaccine;
excretory system; ds.

Homo sapiens.

WO200155313-A2.

02-AUG-2001.

17-JAN-2001; 2001WO-US01323.

31-JAN-2000; 2000US-0179065.
04-FEB-2000; 2000US-0180628.
24-FEB-2000; 2000US-0184664.
02-MAR-2000; 2000US-0186350.
16-MAR-2000; 2000US-0189874.
17-MAR-2000; 2000US-0190076.
18-APR-2000; 2000US-0198123.
19-MAY-2000; 2000US-0205515.
07-JUN-2000; 2000US-0209466.
28-JUN-2000; 2000US-0214886.
30-JUN-2000; 2000US-0215133.
07-JUL-2000; 2000US-0216647.
07-JUL-2000; 2000US-0216880.
11-JUL-2000; 2000US-0217487.
11-JUL-2000; 2000US-0217496.
14-JUL-2000; 2000US-0218290.
26-JUL-2000; 2000US-0220963.
26-JUL-2000; 2000US-0220964.
14-AUG-2000; 2000US-0224518.
14-AUG-2000; 2000US-0224519.
14-AUG-2000; 2000US-0225213.
14-AUG-2000; 2000US-0225214.
14-AUG-2000; 2000US-0225266.
14-AUG-2000; 2000US-0225267.
14-AUG-2000; 2000US-0225270.
14-AUG-2000; 2000US-0225447.
14-AUG-2000; 2000US-0225757.
14-AUG-2000; 2000US-0225758.
14-AUG-2000; 2000US-0225759.
18-AUG-2000; 2000US-0226279.
22-AUG-2000; 2000US-0226581.
22-AUG-2000; 2000US-0226586.
22-AUG-2000; 2000US-0227182.
23-AUG-2000; 2000US-0227009.
30-AUG-2000; 2000US-0228924.
01-SEP-2000; 2000US-0229287.
01-SEP-2000; 2000US-0229343.
01-SEP-2000; 2000US-0229344.
01-SEP-2000; 2000US-0229345.
05-SEP-2000; 2000US-0229509.
05-SEP-2000; 2000US-0229513.
06-SEP-2000; 2000US-0230437.
06-SEP-2000; 2000US-0230438.
08-SEP-2000; 2000US-0231242.
08-SEP-2000; 2000US-0231243.
08-SEP-2000; 2000US-0231244.
08-SEP-2000; 2000US-0231413.
08-SEP-2000; 2000US-0231414.
08-SEP-2000; 2000US-0232080.
08-SEP-2000; 2000US-0232081.
12-SEP-2000; 2000US-0231968.
14-SEP-2000; 2000US-0232397.
14-SEP-2000; 2000US-0232398.
14-SEP-2000; 2000US-0232399.
14-SEP-2000; 2000US-0232400.
14-SEP-2000; 2000US-0232401.
14-SEP-2000; 2000US-0233063.
14-SEP-2000; 2000US-0233064.
14-SEP-2000; 2000US-0233065.
21-SEP-2000; 2000US-0234223.
21-SEP-2000; 2000US-0234274.
25-SEP-2000; 2000US-0234997.
25-SEP-2000; 2000US-0234998.
26-SEP-2000; 2000US-0235484.
27-SEP-2000; 2000US-0235834.
27-SEP-2000; 2000US-0235836.
29-SEP-2000; 2000US-0236327.
29-SEP-2000; 2000US-0236367.
29-SEP-2000; 2000US-0236368.
29-SEP-2000; 2000US-0236369.
29-SEP-2000; 2000US-0236370.

PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239935.
PR 13-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241221.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 01-NOV-2000; 2000US-0246174.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249246.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249267.
PR 17-NOV-2000; 2000US-0249269.
PR 17-NOV-2000; 2000US-0249299.
PR 01-DEC-2000; 2000US-0250160.
PR 01-DEC-2000; 2000US-0250391.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.
PR 05-DEC-2000; 2000US-0256719.
PR 06-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251868.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251889.
PR 11-DEC-2000; 2000US-0254097.
PR 05-JAN-2001; 2001US-0259678.

(HUMA-) HUMAN GENOME SCI INC.
PI Rosen CA, Barash SC, Ruben SM;
XX WPI; 2001-465569/50.
XX Isolated nucleic acid molecule encoding excretory system antigen is
PT used in preventing, treating or ameliorating a medical condition -

XX Example 2; SEQ ID NO 816; 574bp + Sequence Listing; English.
PS The invention relates to novel excretory system related human
XX polynucleotides (AA198567-AA199503) and the encoded proteins
CC (AA199594-AA199913) useful for preventing, treating or ameliorating
CC medical conditions e.g. by protein or gene therapy, especially
CC disorders related to the excretory system. The genes are isolated
CC from a range of human tissues disclosed in the specification. The
CC nucleic acids, proteins, antibodies and (ant)agonists are useful in
CC the diagnosis, treatment and prevention of: (a) cancer, e.g. breast
CC and ovarian cancer and other cancers of the adrenal gland, bone, bone
CC marrow, breast, gastrointestinal tract, liver, lung, or urogenital;
CC (b) immune disorders e.g. Addison's disease, allergies, autoimmune
CC haemolytic anaemia, autoimmune thyroiditis, diabetes mellitus, Crohn's
CC disease, multiple sclerosis, rheumatoid arthritis and ulcerative
CC colitis; (c) cardiovascular disorders such as myocardial ischaemia;
CC (d) wound healing; (e) neurological diseases e.g. cerebral anoxia and
CC epilepsy; and (f) infectious diseases such as viral, bacterial, fungal
CC and parasitic infections.
CC Note: The sequence data for this patent did not form part of the
CC printed specification, but was obtained in electronic format directly
CC from WIPO at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 9839 BP; 2292 A; 2702 C; 2654 G; 2191 T; 0 other;

Query Match 72.7%; Score 1016; DB 22; Length 9839;
Best Local Similarity 99.8%; Pred. No. 0;
Matches 1306; Conservative 0; Mismatches 1; Indels 2; Gaps 2;

QY 62 GAGTACCACTGATCCAGACAGACACATCCCCAGAGAACTACCGCTGCGCCATC 121
DB 8533 GAGTACCACTGATCCAGACAGACACATCCCCAGAGAACTACCGCTGCGCCATC 8592

QY 122 TACCAACAGAGAGAGCTGCTCTTTCAGTTCACACTGAGAGCTGATGCTCT 181
DB 8593 TACCAACAGAGAGAGCTGCTCTTTCAGTTCACACTGAGAGCTGATGCTCT 8652

QY 182 GAGAGCCATGCCAGAGTGGGAGCTTTTGTGTGATCCAAACAGACCACTGAGAGTGAG 241
DB 8653 GAGAGCCATGCCAGAGTGGGAGCTTTTGTGTGATCCAAACAGACCACTGAGAGTGAG 8712

QY 242 CCAAGTGGAGAAAGCCCTTTCAGAGAGAGATGGAGACCTCTGTGAGTTGATAGTACT 301
DB 8713 CCAAGTGGAGAAAGCCCTTTCAGAGAGAGATGGAGACCTCTGTGAGTTGATAGTACT 8772

QY 302 GATCCCCCATCGGAAGTCAAGAGGGGTGCTGAGTGTATAGAGAGATACGTTCT 361
DB 8773 GATCCCCCATCGGAAGTCAAGAGGGGTGCTGAGTGTATAGAGAGATACGTTCT 8832

QY 362 CAAGGAGTCAATTAAGGAGAAATGCTTGGCTCCAGAAAGAAATCCAGCCCTGT 421
DB 8833 CAAGGAGTCAATTAAGGAGAAATGCTTGGCTCCAGAAAGAAATCCAGCCCTGT 8892

QY 422 TACCTCTCACTCTGCCCCCAGAGTGGAGAGTGGTCTTTTCAAGACTGATGAGCA 481
DB 8893 TACCTCTCACTCTGCCCCCAGAGTGGAGAGTGGTCTTTTCAAGACTGATGAGCA 8952

QY 482 AGTGTCCCTGATCCCAACAGACCAATATGTAAGGCTTGTGCTGATCTATGAG 541
DB 8953 AGTGTCCCTGATCCCAACAGACCAATATGTAAGGCTTGTGCTGATCTATGAG 9011

QY 542 GGTGGGCTGAGCAGAGTATCTCAGAGTGGGCTTGGCTGAGAGGAGTACTG 601
DB 9012 GGTGGGCTGAGCAGAGTATCTCAGAGTGGGCTTGGCTGAGAGGAGTACTG 9071

QY 602 CACTGAGCACTGATGATCACTGAGAAACCTCTGAGCAAGAGTAAATCCAGACAG 661
DB 9072 CACTGAGCACTGATGATCACTGAGAAACCTCTGAGCAAGAGTAAATCCAGACAG 9131

QY 662 ACAGATGTGACAGAGCAAAAGTGTCAATATGTAAGTGTAAATGTGAGTTTACAGC 721
DB 9132 ACAGATGTGACAGAGCAAAAGTGTCAATATGTAAGTGTAAATGTGAGTTTACAGC 9191

```
QY 722 CTAGCTATGGAGTCTGCTGCTCTAGTCCAGGATCATGGGGGATGATGCTGCTCTCCAA 781
DB 9192 CTAGCTATGGAGTCTGCTGCTCTAGTCCAGGATCATGGGGGATGATGCTGCTCTCCAA 9251
QY 782 CCGTGTGGGCTGTAGCAAGCTCAGGCTAGTCTCCCACTGGGGGCTGTGCTCCCTCCG 841
DB 9252 CCGTGTGGGCTGTAGCAAGCTCAGGCTAGTCTCCCACTGGGGGCTGTGCTCCCTCCG 9310
QY 842 GGAAGGTTCCGTGGGCGAGCCCATCACTGTGTCAATAGTGTAGAAATGAGTAAAGCC 901
DB 9311 GGAAGGTTCCGTGGGCGAGCCCATCACTGTGTCAATAGTGTAGAAATGAGTAAAGCC 9370
QY 902 CTTGCTGCTGCTGCTGCAATGCCACAGCGGGGCTGGGCTGGGCGAATTCAT 961
DB 9371 CTTGCTGCTGCTGCTGCAATGCCACAGCGGGGCTGGGCTGGGCGAATTCAT 9430
QY 962 CGTGGAGTGTCTCTCAGCTTAAAGTGTGAACAGGAGCTTGGGGGATGCTCCAGAT 1021
DB 9431 CGTGGAGTGTCTCTCAGCTTAAAGTGTGAACAGGAGCTTGGGGGATGCTCCAGAT 9490
QY 1022 GTGGGATATCTGATCCTGGGAGGCTATCTGTGACCTCCCGACAGGGGACACTCCAG 1081
DB 9491 GTGGGATATCTGATCCTGGGAGGCTATCTGTGACCTCCCGACAGGGGACACTCCAG 9550
QY 1082 CCAAGCCAGGGGCTCAGGGGCAAGGTGCAACCTCAGCATGAGCCAAACTGGGGCTCAG 1141
DB 9551 CCAAGCCAGGGGCTCAGGGGCAAGGTGCAACCTCAGCATGAGCCAAACTGGGGCTCAG 9610
QY 1142 GAGAGGTTGTGTGAGGCAAGGACCTGGGGGCGGGGCTGGGGGCTGGGGCTGCTC 1201
DB 9611 GAGAGGTTGTGTGAGGCAAGGACCTGGGGGCGGGGCTGGGGGCTGGGGCTGCTC 9670
QY 1202 ATTTGCTTTCAATAAAGCCCTCAAGAGCCCAAAACAGGCTTCCCTCTCTCAGAT 1261
DB 9671 ATTTGCTTTCAATAAAGCCCTCAAGAGCCCAAAACAGGCTTCCCTCTCTCAGAT 9730
QY 1262 TGAATATCCAGAACTTTTGTACTTCTGTGTGTAATTTGTATTTGTAAATA 1321
DB 9731 TGAATATCCAGAACTTTTGTACTTCTGTGTGTAATTTGTATTTGTAAATA 9790
QY 1322 AAATAAATTTGTAATTAATGATGTTTCACAGCAAACTTCCCTAA 1370
DB 9791 AAATAAATTTGTAATTAATGATGTTTCACAGCAAACTTCCCTAA 9839

RESULT 2
AI63402
ID AI63402 standard; DNA; 9839 BP.
XX
AC AI63402;
AC
DT 22-OCT-2001 (first entry)
XX
DE Human kidney related polynucleotide SEQ ID NO 717.
XX
KW Human; kidney antigen; immunosuppressive; antiarthritic; antineumatic;
KW antiproliferative; cytostatic; cardiant; vasotropic; cerebroprotective;
KW neurotropic; neuroprotective; antibacterial; vitinoid; fungicide;
KW ophthalmological; antiallergic; hepatotropic; antidiabetic;
KW ophthalmological; antiallergic; hepatotropic; antidiabetic;
KW gene therapy; cancer; immune disorder; anticonvulsant; antiparasitic;
KW neurological disease; infection; ds.
XX
XX Homo sapiens.
XX OS
XX PN WO200155323-A2.
XX
XX 02-AUG-2001.
XX
XX 17-JAN-2001; 2001WO-US01343.
XX
XX 31-JAN-2000; 2000US-0179065.
PR
```

```
PR 04-FEB-2000; 2000US-0180628.
PR 24-FEB-2000; 2000US-0184664.
PR 02-MAR-2000; 2000US-0186350.
PR 16-MAR-2000; 2000US-0188874.
PR 17-MAR-2000; 2000US-0190076.
PR 18-APR-2000; 2000US-0198123.
PR 19-MAY-2000; 2000US-0205515.
PR 07-JUN-2000; 2000US-0209467.
PR 28-JUN-2000; 2000US-0214886.
PR 30-JUN-2000; 2000US-0215135.
PR 07-JUL-2000; 2000US-0216647.
PR 07-JUL-2000; 2000US-0216880.
PR 11-JUL-2000; 2000US-0217487.
PR 11-JUL-2000; 2000US-0217496.
PR 14-JUL-2000; 2000US-0218290.
PR 26-JUL-2000; 2000US-0220963.
PR 26-JUL-2000; 2000US-0220964.
PR 14-AUG-2000; 2000US-0224518.
PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0225213.
PR 14-AUG-2000; 2000US-0225214.
PR 14-AUG-2000; 2000US-0225266.
PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225268.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225447.
PR 14-AUG-2000; 2000US-0225757.
PR 14-AUG-2000; 2000US-0225758.
PR 14-AUG-2000; 2000US-0225759.
PR 18-AUG-2000; 2000US-0226279.
PR 22-AUG-2000; 2000US-0226681.
PR 22-AUG-2000; 2000US-0226688.
PR 22-AUG-2000; 2000US-0227182.
PR 23-AUG-2000; 2000US-0227009.
PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 08-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0234984.
PR 26-SEP-2000; 2000US-0235634.
PR 27-SEP-2000; 2000US-0235636.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
```


PR 02-OCT-2000; 2000US-0237038.
 PR 02-OCT-2000; 2000US-0237039.
 PR 02-OCT-2000; 2000US-0237040.
 PR 13-OCT-2000; 2000US-0239935.
 PR 13-OCT-2000; 2000US-0239937.
 PR 20-OCT-2000; 2000US-0240960.
 PR 20-OCT-2000; 2000US-0241221.
 PR 20-OCT-2000; 2000US-0241785.
 PR 20-OCT-2000; 2000US-0241786.
 PR 20-OCT-2000; 2000US-0241787.
 PR 20-OCT-2000; 2000US-0241808.
 PR 20-OCT-2000; 2000US-0241809.
 PR 20-OCT-2000; 2000US-0241826.
 PR 01-NOV-2000; 2000US-024617.
 PR 08-NOV-2000; 2000US-0246474.
 PR 08-NOV-2000; 2000US-0246475.
 PR 08-NOV-2000; 2000US-0246476.
 PR 08-NOV-2000; 2000US-0246477.
 PR 08-NOV-2000; 2000US-0246478.
 PR 08-NOV-2000; 2000US-0246523.
 PR 08-NOV-2000; 2000US-0246524.
 PR 08-NOV-2000; 2000US-0246525.
 PR 08-NOV-2000; 2000US-0246526.
 PR 08-NOV-2000; 2000US-0246527.
 PR 08-NOV-2000; 2000US-0246528.
 PR 08-NOV-2000; 2000US-0246532.
 PR 08-NOV-2000; 2000US-0246609.
 PR 08-NOV-2000; 2000US-0246610.
 PR 08-NOV-2000; 2000US-0246611.
 PR 08-NOV-2000; 2000US-0246613.
 PR 17-NOV-2000; 2000US-0249207.
 PR 17-NOV-2000; 2000US-0249208.
 PR 17-NOV-2000; 2000US-0249209.
 PR 17-NOV-2000; 2000US-0249210.
 PR 17-NOV-2000; 2000US-0249211.
 PR 17-NOV-2000; 2000US-0249212.
 PR 17-NOV-2000; 2000US-0249213.
 PR 17-NOV-2000; 2000US-0249214.
 PR 17-NOV-2000; 2000US-0249215.
 PR 17-NOV-2000; 2000US-0249216.
 PR 17-NOV-2000; 2000US-0249217.
 PR 17-NOV-2000; 2000US-0249218.
 PR 17-NOV-2000; 2000US-0249244.
 PR 17-NOV-2000; 2000US-0249245.
 PR 17-NOV-2000; 2000US-0249264.
 PR 17-NOV-2000; 2000US-0249265.
 PR 17-NOV-2000; 2000US-0249297.
 PR 17-NOV-2000; 2000US-0249299.
 PR 17-NOV-2000; 2000US-0249300.
 PR 01-DEC-2000; 2000US-0250160.
 PR 01-DEC-2000; 2000US-0250391.
 PR 05-DEC-2000; 2000US-0251030.
 PR 05-DEC-2000; 2000US-0251988.
 PR 05-DEC-2000; 2000US-0256719.
 PR 06-DEC-2000; 2000US-0251479.
 PR 08-DEC-2000; 2000US-0251856.
 PR 08-DEC-2000; 2000US-0251868.
 PR 08-DEC-2000; 2000US-0251869.
 PR 08-DEC-2000; 2000US-0251989.
 PR 08-DEC-2000; 2000US-0251990.
 PR 11-DEC-2000; 2000US-0254097.
 PR 05-JAN-2001; 2001US-0259678.
 XX
 PA (HUMA-) HUMAN GENOME SCI INC.
 PI Rosen CA, Barash SC, Ruben SM;
 XX
 XX WPI; 2001-488784/53.
 PT New isolated nucleic acids and polypeptides, useful for diagnosing,
 XX treating and/or preventing human diseases and disorders -
 PS Disclosure; SEQ ID NO 717; 564pp + Sequence Listing; English.

XX The invention relates to novel kidney related polynucleotides
 CC (AA162971-AA163793) and the encoded polypeptides (AA162417-AA164261)
 CC collectively known as kidney antigens and the use of such kidney antigens
 CC for detecting disorders of the kidney, especially kidney cancer and
 CC kidney cancer metastases. The polynucleotides and proteins are also
 CC useful for preventing, treating or ameliorating medical conditions
 CC e.g. by protein or gene therapy. The genes are isolated from a range
 CC of human tissues disclosed in the specification. The nucleic acids,
 CC proteins, antibodies and (antagonists are useful in the diagnosis,
 CC treatment and prevention of: (a) cancer, e.g. breast and ovarian cancer,
 CC and other cancers of the adrenal gland, bone, bone marrow, breast,
 CC gastrointestinal tract, liver, lung, or urogenital; (b) immune disorders
 CC e.g. Addison's disease, allergies, autoimmune hemolytic anaemia,
 CC autoimmune thyroiditis, diabetes mellitus, Crohn's disease, multiple
 CC sclerosis, rheumatoid arthritis and ulcerative colitis;
 CC (c) cardiovascular disorders such as myocardial ischaemia; (d) wound
 CC healing; (e) neurological diseases e.g. cerebral anoxia and epilepsy;
 CC and (f) infectious diseases such as viral, bacterial, fungal and
 CC parasitic infections.
 CC Note: The sequence data for this patent did not form part of the
 CC printed specification, but was obtained in electronic format directly
 CC from WIPO at ftp.wipo.int/pub/published_pcr_sequences.
 XX
 SQ Sequence 9839 BP; 2292 A; 2702 C; 2654 G; 2191 T; 0 other;
 Query Match 72.7%; Score 1016; DB 22; Length 9839;
 Best Local Similarity 99.8%; Pred. No. 0;
 Matches 1306; Conservative 0; Mismatches 1; Indels 2; Gaps 2;
 QY 62 GAGTACCACTGATATCCACAGACACACATCCCGAGAGAGTACCGCTGCGCCATCC 121
 DB 8533 GAGTACCACTGATATCCACAGACACACATCCCGAGAGAGTACCGCTGCGCCATCC 8592
 QY 122 TACACACACGAGAGCTCTCTTTCAGTGTTCACCTGCGAGCTGTGATGTCTGT 181
 DB 8593 TACACACACGAGAGCTCTCTTTCAGTGTTCACCTGCGAGCTGTGATGTCTGT 8652
 QY 182 GAGAGCCATGCCAGAGTGGGCTTTGTGTGACCAACACACACCTGAGAGTGG 241
 DB 8653 GAGAGCCATGCCAGAGTGGGCTTTGTGTGACCAACACACACCTGAGAGTGG 8712
 QY 242 CAGTGGAGAGAGCCCTTTCAGAGAGAGAGAGAGCTCTGAGAGTGTATAGTACT 301
 DB 8713 CAGTGGAGAGAGCCCTTTCAGAGAGAGAGAGAGCTCTGAGAGTGTATAGTACT 8772
 QY 302 GATCCCCATCGAAGTCAAGAGGAGTGTGATGATGAGAGAGTATACGTCTCT 361
 DB 8773 GATCCCCATCGAAGTCAAGAGGAGTGTGATGATGAGAGAGTATACGTCTCT 8832
 QY 362 CAAGGAGTCAATTAAGAGAGAGTGTGCTTTCAGAGAGAGAGAGAGTATACGTCTCT 421
 DB 8833 CAAGGAGTCAATTAAGAGAGAGTGTGCTTTCAGAGAGAGAGAGAGTATACGTCTCT 8892
 QY 422 TACCTCTCACTCTGCCCCCAGAGTGGAGTGTCTTTTCAAGAGCTGATGAGCA 481
 DB 8893 TACCTCTCACTCTGCCCCCAGAGTGGAGTGTCTTTTCAAGAGCTGATGAGCA 8952
 QY 482 AGTGTCTCTGATATCCACAGACACATATGTGAATGCTGTGCTGATCTATCTGAG 541
 DB 8953 AGTGTCTCTGATATCCACAGACACATATGTGAATGCTGTGCTGATCTATCTGAG 9011
 QY 542 GCTGGGCTGACAGAGTATATCTTTCAGAGAGTGGCTTTCAGAGAGAGTATCTG 601
 DB 9012 GCTGGGCTGACAGAGTATATCTTTCAGAGAGTGGCTTTCAGAGAGAGTATCTG 9071
 QY 602 CACTGGACAGACTGATGATCACTGGAGAGCCCTGAGAGAGTATATCTGAGAGAG 661
 DB 9072 CACTGGACAGACTGATGATCACTGGAGAGCCCTGAGAGAGTATATCTGAGAGAG 9131
 QY 662 ACAGATGTGACAGAGACAAAGTGTCAATATATGCTTAAATGTGAGTTTACAGC 721
 DB 9132 ACAGATGTGACAGAGACAAAGTGTCAATATATGCTTAAATGTGAGTTTACAGC 9191

QY 1102 AGAGTGACACCTCAGCATGAGCCAAAGCTGGGGTCAGGAGCAGGTGTGTTGAGCC 1161
 DB 1568 AGAGTGACACCTCAGCATGAGCCAAAGCTGGGGTCAGGAGCAGGTGTGTTGAGCC 1627
 QY 1162 AGAGCTGGGGCGGGGGTGGGGCGGGCTTTCCTCCATTTGTTTCATGAAAGCC 1221
 DB 1628 AGAGCTGGGGCGGGGGTGGGGCGGGCTTTCCTCCATTTGTTTCATGAAAGCC 1687
 QY 1222 TCAGAGCAGCCAAACAGGCTTTCCTCCTCAGATTGAAATCCAGAACTTTTG 1281
 DB 1688 TCAGAGCAGCCAAACAGGCTTTCCTCCTCAGATTGAAATCCAGAACTTTTG 1747
 QY 1282 TACTCTGTGTGTTAAATTTGTTTAAATTAATAATAATAATTAAGTTAATAA 1341
 DB 1748 TACTCTGTGTGTTAAATTTGTTTAAATTAATAATAATAATTAAGTTAATAA 1807
 QY 1342 ATGATGTTTCACAGCAAACTCTTCCCTAATAAAAAAAAA 1380
 DB 1808 ATGATGTTTCACAGCAAACTCTTCCCTAATAAAAAAAAA 1846
 RESULT 4
 AAF44725
 ID AAF44725 standard; cDNA, 2461 BP.
 AC AAF44725;
 XX
 DT 27-MAR-2001 (first entry)
 DE Novel protein kinase cDNA, SEQ ID NO: 106.
 XX
 XX Human; mouse; protein kinase; antiarthritic; antisclerotic; osteopathic;
 XX immunosuppressive; cardiant; renal; antiinflammatory; antiashtmatic;
 XX dermatological; antidiabetic; antifertility; gene therapy; vaccine;
 XX immune disorder; cardiovascular disease; neurodegenerative disease;
 XX cancer; autoimmune disorder; stroke; inflammatory bowel disease;
 XX inflammatory pelvic disease; multiple sclerosis; psoriasis; ss.
 OS Homo sapiens.
 XX
 PN MO200073469-A2.
 XX
 PD 07-DEC-2000.
 PF 26-MAY-2000; 2000MO-US14842.
 XX
 PR 28-MAY-1999; 99US-0136503.
 XX
 XX (SUGEN-) SUGEN INC.
 XX
 PI Plowman GD, Martinez R, Whyte D, Suderanan S.
 XX
 DR WPI; 2001-032161/04.
 XX
 DR P-PSDB; AAB65697.
 XX
 PT Nucleic acids encoding kinase polypeptides, useful for diagnosing and
 PT treating immune-related diseases and disorders, cardiovascular disease,
 PT neurodegenerative diseases and/or cancers -
 XX
 XX Example 1; Fig 2; 310p; English.
 XX
 XX The present sequence encodes a novel protein kinase. The nucleic acids
 XX and the protein kinases they encode may be used in the treatment and
 XX diagnosis of diseases associated with inappropriate kinase expression
 XX such as immune-related diseases and disorders, cardiovascular disease,
 XX neurodegenerative diseases and/or cancers. The nucleic acids and
 XX complementary sequences may also be used as DNA probes in diagnostic
 XX assays. The kinase polypeptides may be used as antigens in the production
 XX of antibodies of kinase expression and activity. Anti-kinase antibodies
 XX and kinase antagonists may also be used to down regulate kinase
 XX expression and activity. Diseases related to kinase expression and
 XX activity include rheumatoid arthritis, atherosclerosis, autoimmune
 XX disorders, complications of organ transplantation, myocardial infarction,

CC immune disorders, cardiomyopathies, strokes, renal failure,
 CC oxidative-stress related disorders, chronic inflammatory bowel disease,
 CC chronic inflammatory pelvic disease, multiple sclerosis, asthma,
 CC osteoarthritis, psoriasis, rhinitis, autoimmunity, diabetes, cancers and
 CC reproductive disorders.
 SQ Sequence 2461 BP; 426 A; 771 C; 811 G; 453 T; 0 other;
 Query Match 62.7%; Score 876; DB 22; Length 2461;
 Best Local Similarity 99.9%; Pred. No. 0;
 Matches 926; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
 QY 442 CAGGTGGAGCTGTGCTTTTCAAGACTGATGAGCCAGCTGCTCCATGCCAACA 501
 DB 1535 CAGGTGGAGCTGTGCTTTTCAAGACTGATGAGCCAGCTGCTCCATGCCAACA 1594
 QY 502 AGACCAATATGTGAAGGCTTGGCTGACCTATCTGAGGGCTGGCTGACCAAGCTGAT 561
 DB 1595 AGACCAATATGTGAAGGCTTGGCTGACCTATCTGAGGGCTGGCTGACCAAGCTGAT 1654
 QY 562 ATCCCTCAGAGCTGGGGCTTGGCTGTGAGGGAGTGACTTGCATGGCAGACTGATGTC 621
 DB 1655 ATCCCTCAGAGCTGGGGCTTGGCTGTGAGGGAGTGACTTGCATGGCAGACTGATGTC 1714
 QY 622 ACCGTGGAAACCCCTGCAGACAAAGCTAACATCCAGACAGAGATGTGACAGACAA 681
 DB 1715 ACCGTGGAAACCCCTGCAGACAAAGCTAACATCCAGACAGAGATGTGACAGACAA 1774
 QY 682 CGTGCAATATGCCAATGTTAAATGTGAGTTTACACAGCTTACCTATGTGGACTGTGC 741
 DB 1775 CGTGCAATATGCCAATGTTAAATGTGAGTTTACACAGCTTACCTATGTGGACTGTGC 1834
 QY 742 TCCCTGTCAGGAATCATGAGGGGTATGACTGCTGCCAACCTGGGGCTGTAAAGCAAG 801
 DB 1835 TCCCTGTCAGGAATCATGAGGGGTATGACTGCTGCCAACCTGGGGCTGTAAAGCAAG 1894
 QY 802 CTCAGGCTAGTCTCCCACTGGGGGCTGTGCCCTTCCCTGAGGACGGTTCCGTGGCAGCC 861
 DB 1895 CTCAGGCTAGTCTCCCACTGGGGGCTGTGCCCTTCCCTGAGGACGGTTCCGTGGCAGCC 1954
 QY 862 CCATCACTGTGTTCAATATGTGAGATGTACTTAAAGCCCTGCTGCTGTGCTGACAA 921
 DB 1955 CCATCACTGTGTTCAATATGTGAGATGTACTTAAAGCCCTGCTGCTGTGCTGACAA 2014
 QY 922 TGCACAGAGGCGGTGGGGGCTGTGGGGGCAATTCATCGTGAAGTGTCTTCAGCT 981
 DB 2015 TGCACAGAGGCGGTGGGGGCTGTGGGGGCAATTCATCGTGAAGTGTCTTCAGCT 2074
 QY 982 TAGGTCTGACAGAGACTTGGCGGGGAGTGTCCAGAGATGTGGGTATTTCTGTAACCTG 1041
 DB 2075 TAGGTCTGACAGAGACTTGGCGGGGAGTGTCCAGAGATGTGGGTATTTCTGTAACCTG 2134
 QY 1042 GGAGGCTATCTGTGACCTCCGACAGGGGACACTCCAGGCCAGCCAGGGGTCAAGGGCT 1101
 DB 2135 GGAGGCTATCTGTGACCTCCGACAGGGGACACTCCAGGCCAGCCAGGGGTCAAGGGCT 2194
 QY 1102 AAGAGTGACACCTCAGCATGAGCCAAAGCTGGGGTCAGGAGCAGGTGTGTTGAGCC 1161
 DB 2195 AAGAGTGACACCTCAGCATGAGCCAAAGCTGGGGTCAGGAGCAGGTGTGTTGAGCC 2254
 QY 1162 AGAGCTGGGGCGGGGGTGGGGCGGGCTTTCCTCCTCAGATTGAAATCCAGAAAGCC 1221
 DB 2255 AGAGCTGGGGCGGGGGTGGGGCGGGCTTTCCTCCTCAGATTGAAATCCAGAAAGCC 2314
 QY 1222 TCAGAGCAGCCAAACAGGCTTTCCTCCTCAGATTGAAATCCAGAAAGCC 1281
 DB 2315 TCAGAGCAGCCAAACAGGCTTTCCTCCTCAGATTGAAATCCAGAAAGCC 2374
 QY 1282 TACTCTGTGTGTTAAATTTGTTTAAATTAATAATAATAATTAAGTTAATAA 1341
 DB 2375 TACTCTGTGTGTTAAATTTGTTTAAATTAATAATAATAATTAAGTTAATAA 2434
 QY 1342 ATGATGTTTCACAGCAAACTCTTCCCT 1368

Db 2435 ATGATGTTTCACGACAACTCTTCCT 2461

RESULT 5

ABN94640/C ID ABN94640 standard; DNA; 427 BP.

AC ABN94640;

DT 13-AUG-2002 (first entry)

DE Gene #1138 used to diagnose liver cancer.

XX Gene; liver cancer; db; hepatocellular carcinoma; hepatocytic;
 KW metastatic liver tumor; cytotoxic; expression profile; disease state;
 KM disease progression; drug toxicity; drug efficacy; drug metabolism.

XX Homo sapiens.

OS WO200229103-A2.

PN 11-APR-2002.

PD 02-OCT-2001; 2001MO-US30589.

PR 02-OCT-2000; 2000US-237054P.

XX (GENE-) GENE LOGIC INC.

PI Horne D, Alvares C, Peres-Da-Silva S, Vockley JG;

DR WPI; 2002-426119/45.

XX Diagnosing and detecting the progression of liver cancer,

PT hepatocellular carcinoma or metastatic liver tumor in a patient,

PT involves detecting the level of expression of two or more genes in a

PT liver tissue sample

XX Claim 1; SEQ ID NO 1138; 298bp; English.

XX The invention relates to a novel method for diagnosing and detecting the

CC progression of liver cancer, hepatocellular carcinoma or metastatic liver

CC tumour in a patient, and differentiating metastatic liver cancer from

CC hepatocellular carcinoma in a patient, involving detecting the level of

CC expression of two or more genes represented in ABN93503-ABN97455 in a

CC tissue sample. The method of the invention has hepatocytic, and

CC cytosolic activity. The method is useful for diagnosing and detecting

CC the progression of liver cancer, hepatocellular carcinoma and metastatic

CC liver carcinoma in a patient. The method is useful for identifying

CC expression profiles which serve as useful diagnostic markers as well as

CC markers that can be used to monitor disease states, disease progression,

CC drug toxicity, drug efficacy and drug metabolism.

CC Note: The sequence data for this patent did not form part of the printed

CC specification, but was obtained in electronic format directly from WIPO

CC at ftp.wipo.int/pub/published_pct_sequences.

XX SQ Sequence 427 BP; 107 A; 126 C; 92 G; 102 T; 0 other;

XX Query Match 23.0%; Score 322; DB 24; Length 427;

XX Best Local Similarity 99.5%; Pred. No. 2.4e-115;

XX Matches 422; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 936 GTGGGGGCTGCTGGGGAACAATCCATCGTGGAGTGTCTCTCACTTAGTCTGGACAG 995

DB 427 GTGGGGGCTGCTGGGGAACAATCCATCGTGGAGTGTCTCTCACTTAGTCTGGACAG 368

QY 996 AGACTTGGCGGGGATGCTCCAGATGTTGGTATTTCTGTACCTGGGAGGCTATCTCTG 1055

DB 367 AGACTTGGCGGGGATGCTCCAGATGTTGGTATTTCTGTACCTGGGAGGCTATCTCTG 308

QY 1056 ACCGCCGACAGGGAGCACTCCAGGCCAGCCAGGGGTCAGGGGCGAGAGTGACACACT 1115

Db 307 ACCTCCGACAGGGAGCACTCCAGGCCAGCCAGGGGTCAGGGGCGAGAGTGACACACT 248

QY 1116 CAGCATGAGCCAAACACTGGGCTCAGGAGCAGGAGTGTGTTGAGCCAGACCTGGGGCGG 1175

Db 247 CAGCATGAGCCAAACACTGGGCTCAGGAGCAGGAGTGTGTTGAGCCAGACCTGGGGCGG 188

QY 1176 GGGTGGGGCCGGGGCTTTTCCTTCAGTTGCAATTCCTTTTCATGAAAGCCTCAAGACCCAAA 1235

Db 187 GGGTGGAGACCGGGGCTTTTCCTTCAGTTGCAATTCCTTTTCATGAAAGCCTCAAGACCCAAA 128

QY 1236 ACCAGGCTTTCCCTTCCTTCAGTTGCAATTCCTTTTCATGAAAGCCTTTTGTGTGT 1295

Db 127 ACCAGGCTTTCCCTTCCTTCAGTTGCAATTCCTTTTCATGAAAGCCTTTTGTGTGT 68

QY 1236 TAAATGTTTATTTTGTAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 1355

Db 67 TAAATGTTTATTTTGTAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 8

QY 1356 CAAA 1359

Db 7 CAAA 4

RESULT 6

AAH86413/C ID AAH86413 standard; DNA; 150 BP.

XX AAH86413;

XX 27-FEB-2002 (first entry)

DE Human single nucleotide polymorphism containing DNA sequence #1270.

XX Biallelic marker; polymorphism; human; disease; diagnosis; treatment;

XX phenotypic trait; gene therapy; forensic; paternity; mapping; cancer;

XX transgenic; single nucleotide polymorphism; SNP; ss.

XX Homo sapiens.

XX Key Location/Qualifiers

XX Variation replace(25,T)

XX FT /tag= a /standard_name= "single nucleotide polymorphism"

XX PN MO9953095-A2.

XX PD 21-OCT-1999.

XX PF 30-MAR-1999; 99WO-US06893.

XX PR 09-APR-1998; 98US-0057871.

XX PA (WHED) WHITEHEAD INST BIOMEDICAL RES.

XX PI Lander ES, Wang D, Hudson T;

XX DR WPI; 1999-620443/53.

XX PT Polymorphic human genomic sequences and related allele-specific probes

XX of disease

XX Claim 1; Page 167; 330pp; English.

XX This invention describes novel human nucleic acid segments (1)

CC containing polymorphic sites. The polymucleotides of (1) are used for,

CC e.g. correlating disease polymorphisms (or disease susceptibility) or

CC other phenotypic traits (e.g. baldness, obesity, fertility, strength,

CC response to drugs etc.); diagnosing and monitoring e.g. cancer,

CC inflammation, heart or central nervous system diseases; detecting

CC susceptibility to microbial infection; treating or preventing such

CC diseases; forensic analysis; gene therapy; paternity testing; mapping

CC genomic loci associated with phenotypic traits (and subsequent cloning

CC of the genes responsible); and the production of transgenic organisms.
CC Antibodies raised against (I) are useful as diagnostic and therapeutic
CC tools and in drug screening. AAH85144 - AAH87644 represent the human
CC DNA sequences containing biallelic polymorphic sites described in the
CC invention.

XX Sequence 150 BP; 52 A; 19 C; 30 G; 48 T; 1 other;

Query Match 8.1%; Score 113; DB 20; Length 150;
Best Local Similarity 100.0%; Pred. No. 2.8e-34;
Matches 113; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1256 CGAGTTGAAATCCAGAACTCTTTGACTCTTGTGGTTAAATGTTATTTTGTAA 1315

DB 113 CGAGTTGAAATCCAGAACTCTTTGACTCTTGTGGTTAAATGTTATTTTGTAA 54

QY 1316 AAAATTAATAATTAAGTTAATAAATGATGTTTACAGCAAACTCTTCCT 1368

DB 53 AAAATTAATAATTAAGTTAATAAATGATGTTTACAGCAAACTCTTCCT 1

RESULT 7

AAH86414/C
ID AAH86414 standard; DNA; 150 BP.

XX AAH86414;

XX 27-FEB-2002 (first entry)

DE Human single nucleotide polymorphism containing DNA sequence #1271.

XX Biallelic marker; polymorphism; human; disease; diagnosis; treatment;

KW phenotypic trait; gene therapy; forensic; paternity; mapping; cancer;

KW transgenic; single nucleotide polymorphism; SNP; ss.

XX Homo sapiens.

OS Location/Qualifiers

FT Key replacement(92,T) /tag= a

FT Variation /standard_name= "single nucleotide polymorphism"

XX WO953095-A2.

XX 21-OCT-1999.

XX 30-MAR-1999; 99WO-US06893.

XX 09-APR-1998; 98US-0057871.

XX (WHED) WHITEHEAD INST BIOMEDICAL RES.

XX Lander ES, Wang D, Hudson T;

XX WPI; 1999-620443/53.

PT Polymorphic human genomic sequences and related allele-specific probes

PT of disease - useful for genetic analysis, e.g. diagnosis and monitoring

XX Claim 1; Page 167; 330pp; English.

XX This invention describes novel human nucleic acid segments (I) containing polymorphic sites. The polymorphic sites of (I) are used for, e.g. correlating disease polymorphisms (or disease susceptibility) or other phenotypic traits (e.g. baldness, obesity, fertility, strength, response to drugs etc.); diagnosing and monitoring e.g. cancer, inflammation, heart or central nervous system diseases; detecting susceptibility to microbial infection; treating or preventing such diseases; forensic analysis; gene therapy; paternity testing; mapping genomic loci associated with phenotypic traits and subsequent cloning of the genes responsible; and the production of transgenic organisms. Antibodies raised against (I) are useful as diagnostic and therapeutic

CC tools and in drug screening. AAH85144 - AAH87644 represent the human
CC DNA sequences containing biallelic polymorphic sites described in the
CC invention.

XX Sequence 150 BP; 52 A; 19 C; 30 G; 48 T; 1 other;

Query Match 8.1%; Score 113; DB 20; Length 150;
Best Local Similarity 100.0%; Pred. No. 2.8e-34;
Matches 113; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1256 CGAGTTGAAATCCAGAACTCTTTGACTCTTGTGGTTAAATGTTATTTTGTAA 1315

DB 113 CGAGTTGAAATCCAGAACTCTTTGACTCTTGTGGTTAAATGTTATTTTGTAA 54

QY 1316 AAAATTAATAATTAAGTTAATAAATGATGTTTACAGCAAACTCTTCCT 1368

DB 53 AAAATTAATAATTAAGTTAATAAATGATGTTTACAGCAAACTCTTCCT 1

RESULT 8

AAH87838
ID AAH87838 standard; cDNA; 1837 BP.

XX AAH87838;

XX 13-FEB-2002 (first entry)

DE DNA encoding novel human diagnostic protein #14642.

XX Human; chromosome mapping; gene mapping; gene therapy; forensic;

KW food supplement; medical imaging; diagnostic; genetic disorder; ss.

XX Homo sapiens.

XX WO200175067-A2.

XX 11-OCT-2001.

XX 30-MAR-2001; 2001WO-US08631.

XX 31-MAR-2000; 2000US-0540217.

XX 23-AUG-2000; 2000US-0649167.

XX (HYSE-) HYSEQ INC.

XX Drmanac RT, Liu C, Tang YT;

XX WPI; 2001-639362/73.

XX P-PSDB; ABG14651.

PT New isolated polynucleotide and encoded polypeptides, useful in

PT diagnostics, forensics, gene mapping, identification of mutations

PT responsible for genetic disorders or other traits and to assess

PT biodiversity

XX Claim 1; SEQ ID NO 14642; 103pp; English.

XX The invention relates to isolated polynucleotide (I) and polypeptide (II) sequences. (I) is useful as hybridisation probes, CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome CC and gene mapping, and in recombinant production of (II). The CC polynucleotides are also used in diagnostics as expressed sequence tags CC for identifying expressed genes. (I) is useful in gene therapy techniques CC to restore normal activity of (II) or to treat disease states involving CC (II). (II) is useful for generating antibodies against it, detecting or CC quantitating a polypeptide in tissue, as molecular weight markers and as CC a food supplement. (II) and its binding partners are useful in medical CC imaging of sites expressing (II). (I) and (II) are useful for treating CC disorders involving aberrant protein expression or biological activity. CC The polypeptide and polynucleotide sequences have applications in CC diagnostics, forensics, gene mapping, identification of mutations CC responsible for genetic disorders or other traits to assess biodiversity CC and to produce other types of data and products dependent on DNA and

CC amino acid sequences. AAS64197-AAS94564 represent novel human
CC diagnostic coding sequences of the invention.
CC Note: The sequence data for this patent did not appear in the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
CC

SO Sequence 1837 BP; 424 A; 540 C; 471 G; 402 T; 0 other;

Query Match 4.9%; Score 69; DB 23; Length 1837;
Best Local Similarity 100.0%; Pred. No. 2.4e-17;
Matches 69; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 28 GTATCGAGACTCCAGCGGCAAGCGAGTACCGAGTACCGAGTACCGAGAGCAGC 87
DB 1033 GTATCGAGACTCCAGCGGCAAGCGAGTACCGAGTACCGAGTACCGAGAGCAGC 1092

QY 88 CATCCCCCA 96
DB 1093 CATCCCCCA 1101

RESULT 9
AAS78836
ID AAS78836 standard; cDNA; 370 BP.

AC AAS78836;

DT 13-FEB-2002 (first entry)

DE DNA encoding novel human diagnostic protein #14640.

KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW food supplement; medical imaging; diagnostic; genetic disorder; ss.

OS Homo sapiens.

PN WO200175067-A2.

XX 11-OCT-2001.

PF 30-MAR-2001; 2001WO-US08631.

PR 31-MAR-2000; 2000US-0540217.

PR 23-AUG-2000; 2000US-0649167.

PA (HYSR-) HYSRQ INC.

PI Drmanac RT, Liu C, Tang YT;

DR WPI: 2001-639362/73.

DR P-PSDB; AABG14649.

PT New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity

PS Claim 1; SEQ ID No 14640; 103bp; English.

XX The invention relates to isolated polynucleotide (I) and
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC and gene mapping, and in recombinant production of (II). The
CC polynucleotides are also used in diagnostics as expressed sequence tags
CC for identifying expressed genes. (I) is useful in gene therapy techniques
CC to restore normal activity of (II) or to treat disease states involving
CC (II). (II) is useful for generating antibodies against it, detecting or
CC quantitating a polypeptide in tissue, as molecular weight markers and as
CC a food supplement. (II) and its binding partners are useful in medical
CC imaging of sites expressing (II). (I) and (II) are useful for treating
CC disorders involving aberrant protein expression or biological activity.
CC The polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations

CC responsible for genetic disorders or other traits to assess biodiversity
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. AAS64197-AAS94564 represent novel human
CC diagnostic coding sequences of the invention.
CC Note: The sequence data for this patent did not appear in the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
CC

SO Sequence 370 BP; 67 A; 126 C; 101 G; 76 T; 0 other;

Query Match 3.8%; Score 53; DB 23; Length 370;
Best Local Similarity 100.0%; Pred. No. 4.5e-11;
Matches 53; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 95 CAGGAGACTACCGGCTGCGCTACCTACCGAGGAGGAGGCTGCTCTTC 147
DB 317 CAGGAGACTACCGGCTGCGCTACCTACCGAGGAGGAGGCTGCTCTTC 369

RESULT 10
AAB44726
ID AAB44726 standard; cDNA; 1886 BP.

AC AAB44726;

DT 27-MAR-2001 (first entry)

DE Novel protein kinase cDNA, SEQ ID NO: 107.

KW Human; mouse; protein kinase; antiarthritic; antisclerotic; osteopathic;
KW immunosuppressive; cardiac; renal; antiinflammatory; antiasthmatic;
KW dermatological; antidiabetic; antifertility; gene therapy; vaccine;
KW immune disorder; cardiovascular disease; neurodegenerative disease;
KW cancer; autoimmune disorder; stroke; inflammatory bowel disease;
KW inflammatory pelvic disease; multiple sclerosis; psoriasis; ss.

OS Mus musculus.

PN WO200073469-A2.

XX 07-DEC-2000.

PF 26-MAY-2000; 2000WO-US14842.

PR 28-MAY-1999; 99US-0136503.

PA (SUGR-) SUGEN INC.

PI Plowman GD, Martinez R, Whyte D, Sudersanam S;

DR WPI: 2001-032161/04.

DR P-PSDB; AAB59576.

PT Nucleic acids encoding kinase polypeptides, useful for diagnosing and
PT treating immune-related diseases and disorders, cardiovascular disease,
PT neurodegenerative diseases and/or cancers

PS Disclosure; Fig 2; 310bp; English.

XX The present sequence encodes a novel protein kinase. The nucleic acids
CC and the protein kinases they encode may be used in the treatment and
CC diagnosis of diseases associated with inappropriate kinase expression
CC such as immune-related diseases and disorders, cardiovascular disease,
CC neurodegenerative diseases and/or cancers. The nucleic acids and
CC complementary sequences may also be used as DNA probes in diagnostic
CC assays. The kinase polypeptides may be used as antigens in the production
CC of antibodies of kinase expression and activity. Anti-kinase antibodies
CC and kinase antagonists may also be used to down regulate kinase
CC expression and activity. Diseases related to kinase expression and
CC activity include Rheumatoid arthritis, atherosclerosis, autoimmune
CC disorders, complications of organ transplantation, myocardial infarction,
CC immune disorders, cardiomyopathies, strokes, renal failure,
CC oxidative-stress related disorders, chronic inflammatory bowel disease,

CC chronic inflammatory pelvic disease, multiple sclerosis, asthma,
CC osteoarthritis, psoriasis, rhinitis, autoimmunity, diabetes, cancers and
CC reproductive disorders.
XX
XX Sequence 1886 BP; 384 A; 559 C; 553 G; 390 T; 0 other;
Query Match 2.6%; Score 36; DB 22; Length 1886;
Best Local Similarity 100.0%; Pred. No. 0.00015;
Matches 36; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 203 GCCTTGTGTGTCACCAACGACCACTGTGACAGGT 238
DB 981 GCCTTGTGTGTCACCAACGACCACTGTGACAGGT 1016
RESULT 11
AB085200/c
ID AB085200 standard; DNA; 386 BP.
XX
AC AB085200;
XX
DT 05-SEP-2002 (first entry)
XX
DE Arabidopsis thaliana expressed polynucleotide SEQ ID NO 70.
XX
KM Plant; Arabidopsis; transgenic; fungicide; insecticide; ds.
XX
OS Arabidopsis thaliana.
XX
PN US2002062014-A1.
XX
PD 23-MAY-2002.
XX
PF 26-JUN-2001; 2001US-0770791.
XX
PR 27-JUN-2000; 2000US-178480P.
XX
XX (GORL/) GORLACH J.
PA (ANY/) AN Y.
PA (HAM/) HAMILTON C M.
PA (PRIC/) PRICE J L.
PA (RAIN/) RAINES T M.
PA (YUY/) YU Y.
PA (RAME/) RAMEAKA J G.
PA (PAGE/) PAGE A.
PA (MATH/) MATTHEW A V.
PA (LEDF/) LEDFORD B L.
PA (WOES/) WOESSNER J P.
PA (HAAS/) HAAS W D.
PA (GARC/) GARCIA C A.
PA (KRIC/) KRICKER M.
PA (SLAT/) SLATER T.
PA (DAVI/) DAVIS K R.
PA (ALLE/) ALLEN K.
PA (HOFF/) HOFFMAN N.
PA (HURB/) HURBAN P.
XX
PI Gorlach J, An Y, Hamilton CM, Price JL, Raines TM, Yu Y,
PI Rameka JG, Page A, Mathew AV, Ledford BL, Woessner JP, Haas WD,
PI Garcia CA, Kricker M, Slater T, Davis KR, Allen K, Hoffman N,
PI Hurban P;
XX
XX WPI; 2002-479265/51.
XX
XX New nucleic acid sequences of Arabidopsis thaliana and their encoded
XX products are useful to produce transgenic plants, to screen for
XX biologically active agents such as fungicides and insecticides and in
XX genetic studies -
XX
XX Claim 1; SEQ ID NO 70; 18pp + Sequence Listing; English.
XX
XX The invention relates to a novel nucleic acid of Arabidopsis thaliana
XX comprising a sequence capable of hybridising under stringency to one of

CC the 999 sequences referred to but not defined in the specification
CC (AB085131-AB086129). The nucleic acid sequences are useful to identify
CC homologous or related genes, to produce compositions that modulate
CC expression or function of the encoded protein, to map functional regions
CC of the protein, to study associated physiological pathways, to
CC genetically manipulate cells and plants. The encoded products are useful
CC to screen for biologically active agents such as fungicides or
CC insecticides and to elucidate biochemical pathways.
XX
XX Sequence 386 BP; 116 A; 76 C; 75 G; 119 T; 0 other;
Query Match 2.1%; Score 30; DB 24; Length 386;
Best Local Similarity 100.0%; Pred. No. 0.037;
Matches 30; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1361 TCTTCCTTAAAAA 1390
DB 31 TCTTCCTTAAAAA 2
RESULT 12
ABL80663/c
ID ABL80663 standard; cDNA; 421 BP.
XX
AC ABL80663;
XX
DT 17-MAY-2002 (first entry)
XX
DE Human ovarian cancer related cDNA clone SEQ ID NO:3641.
XX
KM Human; ovarian cancer; ovarian tumour; cytostatic; gene; ss.
XX
OS Homo sapiens.
XX
PN WO200192581-A2.
XX
PD 06-DEC-2001.
XX
PF 29-MAY-2001; 2001WO-US17756.
XX
PR 26-MAY-2000; 2000US-207484P.
XX
XX (CORI-) CORIXA CORP.
PA
PI Algate PA, Harlocker SL, Jones R;
XX
XX WPI; 2002-122075/16.
XX
XX Composition for therapy and diagnosis of ovarian cancer comprising
XX polypeptide of a ovarian tumor polypeptide, polynucleotide encoding
XX polypeptide, antibody specific to polypeptide or T cell expressing
XX polypeptide -
XX
XX Claim 1; SEQ ID 3641; 489pp; English.
XX
XX The present invention describes a composition (I) comprising: carriers
XX and immunostimulants; and a polypeptide (II) of a ovarian tumour
XX polypeptide encoded by a polynucleotide (III) having a cDNA sequence
XX (S1) from the 10912 nucleotide sequences as given in ABL77023 to
XX ABL87934, (III) encoding (II) having a sequence (S2), a T cell
XX population of (II), or antigen presenting cells that express (II).
XX (I) has cytostatic activity. An oligonucleotide (IV) that hybridises to
XX (S1) can be used for detecting ovarian cancer in a patient's biological
XX sample preferably serum or ovarian tissue. The method comprises
XX contacting a biological sample from a patient with (IV), detecting the
XX amount of polynucleotide hybridising to (IV) and comparing the amount to
XX a predetermined cutoff value and thereby detecting ovarian cancer in the
XX patient, where the amount of polynucleotide hybridising to (IV) is
XX detected preferably by polymerase chain reaction (PCR). (I) comprising
XX (III) and/or (II) is useful for stimulating and/or expanding T cells
XX specific for an ovarian tumour protein comprising contacting T cells
XX with (III) or (II). (III) is useful in design and preparation of
XX ribozyme molecules for inhibiting expression of the tumour polypeptides

CC and proteins in tumour cells; and to isolate a full length gene from a
CC suitable library e.g., a tumour cDNA library using well known
CC techniques.
XX
SQ Sequence 421 BP; 113 A; 77 C; 100 G; 131 T; 0 other;
Query Match 2.1%; Score 30; DB 24; Length 421;
Best Local Similarity 100.0%; Pred. No. 0.037;
Matches 30; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1361 TCTTCCCTAAAAA 1390
Db 30 TCTTCCCTAAAAA 1
RESULT 13
ABN99116/c
ID ABN99116 standard; DNA; 777 BP.
AC ABN99116;
XX
DT 01-AUG-2002 (first entry)
XX
DE Arabidopsis thaliana expressed polynucleotide SEQ ID NO 884.
XX
XX Arabidopsis thaliana, plant; insecticide; fungicide; transgenic; stress;
XX disease; crop; thale cress; tolerance factor; insect; pathogen;
XX nutrition; ds.
XX
OS Arabidopsis thaliana.
XX
PN US2002023281-A1.
XX
PD 21-FEB-2002.
XX
PF 26-JAN-2001; 2001US-0770445.
XX
PR 27-JAN-2000; 2000US-178472P.
XX
XX (GORL/) GORLACH J.
XX (ANY/) AN Y.
XX (HAM/) HAMILTON C M.
XX (PRIC/) PRICE J L.
XX (RAIN/) RAINES T M.
XX (YUY/) YU Y.
XX (RAME/) RAMEAKA J G.
XX (PAGE/) PAGE A.
XX (MATH/) MATHEW A V.
XX (LEDF/) LEDFORD B L.
XX (WOES/) WOESSNER J P.
XX (HAAS/) HAAS W D.
XX (GARC/) GARCIA C A.
XX (KRIC/) KRICKER M.
XX (SLAT/) SLATER T.
XX (DAVI/) DAVIS K R.
XX (ALIE/) ALLEN K.
XX (HOFF/) HOFFMAN N.
XX (HURE/) HUREN P.
XX
XX Gorlach J, An Y, Hamilton CM, Price JL, Raines TM, Yu Y,
XX Rameaka JG, Page A, Mathew AV, Ledford BL, Woessner JP, Haas WD;
XX Garcia CA, Kricker M, Slater T, Davis KR, Allen K, Hoffman N;
XX Hurean P;
XX
XX WPI; 2002-400781/43.
XX
XX
XX New Arabidopsis thaliana nucleic acid for identifying homologous genes,
XX producing compositions that modulate the expression or function of its
XX encoded protein, and mapping functional regions of protein
XX
XX Claim 1; SEQ ID NO 884; 49bp + Sequence Listing; English.
XX
XX The invention relates to an Arabidopsis thaliana nucleic acid (I)

CC comprising a sequence capable of hybridising under stringent conditions
CC to a sequence selected from any one of 999 sequences (ABN98233-ABN99231),
CC given in the specification or its fragment. A polypeptide (II) encoded by
CC (I), a transgenic plant (III) comprising an exogenous nucleic acid or a
CC genetically modified cell (IV) comprising an exogenous nucleic acid, is
CC useful for screening a candidate agent for its biological effect. (I) is
CC useful in identifying homologous or related genes, in producing
CC compositions that modulate the expression or function of its encoded
CC protein, mapping functional regions of the protein and in studying
CC associated physiological pathways. (I) is also useful for the genetic
CC manipulation of cells, particularly plant cells. (I) is also useful in
CC screening assays of various plant strains to determine the strains that
CC are best capable of withstanding a particular disease or environmental
CC stress. (II) and (III) are useful for screening of biologically active
CC agents, e.g. fungicides, insecticides, etc., for elucidating biochemical
CC pathways. The screened agents are useful in improved methods of treating
CC crops to prevent or treat disease. (II) are also useful in screening
CC programs to identify agents that mimic or enhance the action of tolerance
CC factors. Such agents are useful in improved methods of treating crops to
CC enhance their tolerance to environmental stress. (I) is also useful
CC for enhancing or inhibiting production of a biosynthetic product in a
CC plant. (III) is useful for identifying other mediators that may induce
CC expression of proteins of interest, for establishing the extent to which
CC any specific insect and/or pathogen is responsible for damage to a
CC particular plant, for identifying other mediators that enhance or induce
CC tolerance to environmental stress, for identifying factors involved in
CC biosynthetic pathways of nutritional, commercial, or medicinal value and
CC for identifying products of nutritional, commercial or medicinal
CC value. (IV) is useful in the study of genetic function and regulation,
CC for alteration of the cellular metabolism and for screening compounds
CC that may affect the biological function of the gene or gene products.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from USPRO
CC at seqdata.uspto.gov/sequence.html?DocID=999909770445.
XX
SQ Sequence 777 BP; 192 A; 196 C; 149 G; 240 T; 0 other;
Query Match 2.1%; Score 30; DB 24; Length 777;
Best Local Similarity 100.0%; Pred. No. 0.034;
Matches 30; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1361 TCTTCCCTAAAAA 1390
Db 45 TCTTCCCTAAAAA 16
RESULT 14
AAS46209
ID AAS46209 standard; cDNA; 2457 BP.
XX
AC AAS46209;
XX
DT 18-DEC-2001 (first entry)
XX
DE Human DNA encoding PRO polypeptide sequence #285.
XX
XX PRO polypeptide; mammal; tumour; cancer; human; cattle; horse; sheep; ss;
XX dog; cat; pig; goat; rabbit; tumour necrosis factor alpha; TNF-alpha;
XX blood; chondrocyte cell; cell proliferation; cell differentiation; colon;
XX adrenal; lung; breast; prostate; rectum; cervix; liver; genetic disorder;
XX PCR primer.
XX
XX Homo sapiens.
XX
XX
XX WO200168848-A2.
XX
XX
XX 20-SEP-2001.
XX
XX
XX 28-FEB-2001; 2001WO-US06520.
XX
XX
XX 01-MAR-2000; 2000WO-US05601.
XX
XX 02-MAR-2000; 2000WO-US05841.
XX
XX 03-MAR-2000; 2000US-187202P.

PR 06-MAR-2000; 2000US-186968P.
 PR 14-MAR-2000; 2000US-189320P.
 PR 14-MAR-2000; 2000US-189328P.
 PR 15-MAR-2000; 2000MO-US06884.
 PR 21-MAR-2000; 2000US-190828P.
 PR 21-MAR-2000; 2000US-191007P.
 PR 21-MAR-2000; 2000US-191048P.
 PR 21-MAR-2000; 2000US-191314P.
 PR 28-MAR-2000; 2000US-192655P.
 PR 29-MAR-2000; 2000US-193032P.
 PR 29-MAR-2000; 2000US-193053P.
 PR 30-MAR-2000; 2000MO-US08439.
 PR 04-APR-2000; 2000US-194449P.
 PR 04-APR-2000; 2000US-194647P.
 PR 11-APR-2000; 2000US-195975P.
 PR 11-APR-2000; 2000US-196000P.
 PR 11-APR-2000; 2000US-196187P.
 PR 11-APR-2000; 2000US-196690P.
 PR 11-APR-2000; 2000US-196820P.
 PR 18-APR-2000; 2000US-198121P.
 PR 18-APR-2000; 2000US-198585P.
 PR 25-APR-2000; 2000US-199397P.
 PR 25-APR-2000; 2000US-199550P.
 PR 25-APR-2000; 2000US-199654P.
 PR 03-MAY-2000; 2000US-201516P.
 PR 17-MAY-2000; 2000MO-US13705.
 PR 22-MAY-2000; 2000MO-US14042.
 PR 30-MAY-2000; 2000MO-US14941.
 PR 02-JUN-2000; 2000MO-US15264.
 PR 05-JUN-2000; 2000US-209832P.
 PR 28-JUL-2000; 2000MO-US20710.
 PR 22-AUG-2000; 2000US-064848.
 PR 24-AUG-2000; 2000MO-US23328.
 PR 08-NOV-2000; 2000MO-US30952.
 PR 01-DEC-2000; 2000MO-US32578.
 PR 20-DEC-2000; 2000MO-US34556.
 XX
 XX (GENTH) GENENTECH INC.
 PI Baker KP, Chen J, Desnoyers L, Goddard A, Godowski PJ, Gurney AL;
 PI Pan J, Smith V, Watanabe CK, Wood WT, Zhang Z;
 DR P-PSDB; AA0293308.
 XX
 XX Novel nucleic acids encoding PRO polypeptides, used to diagnose the
 PT presence of tumours, such as prostate and breast tumours, in mammals and
 PT to screen for modulators of the compounds -
 XX
 XX Claim 2; Fig 569; 774pp; English.
 XX
 CC Sequences AAS45925-AA46231 represent DNA molecules encoding and PCR
 CC primers for PRO polypeptides of the invention. The sequences of the
 CC invention can be used to detect the presence of a tumour in a mammal by
 CC comparing the level of expression of a PRO polypeptide in a test sample
 CC of cells from the animal and a control sample of normal cells, whereby a
 CC higher level of expression in the test sample indicates the presence of a
 CC tumour in the mammal. Mammals include dogs, cats, cattle, horses, sheep,
 CC pigs, goats and rabbits but are preferably human. The polypeptides can be
 CC used to stimulate tumour necrosis factor (TNF) alpha release from human
 CC blood, when contacted with it. A specific polypeptide can be used to
 CC stimulate the proliferation or differentiation of chondrocyte cells. The
 CC PRO proteins can be used to determine the presence of tumours and also
 CC susceptibility to tumour development, particularly adrenal, lung, colon,
 CC breast, prostate, rectal, cervical, or liver tumours, in mammalian
 CC subjects. The oligonucleotide probes specific for the PRO nucleic acids
 CC can be used for genetic analysis of individuals with genetic disorders.
 XX
 XX SQ Sequence 2457 BP; 678 A; 629 C; 568 G; 582 T; 0 other;

Query Match 2.1%; Score 30; DB 22; Length 2457;
 Best Local Similarity 100.0%; Pred. No. 0.03;
 Matches 30; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1361 TCTTCCTAAAAAAAAAAAAAAAAAAAAA 1390
 |||||
 DB 2423 TCTTCCTAAAAAAAAAAAAAAAAAAAAA 2452
 |||||
 RESULT 15
 AAD31150/c
 ID AAD31150 standard; cDNA; 3239 BP.
 XX
 AC AAD31150;
 DT 27-AUG-2002 (first entry)
 XX
 DE Human carbohydrate-associated protein (CARBAP) -2 cDNA.
 XX
 KW Human; carbohydrate-associated protein-2; CARBAP-2; arteriosclerosis;
 KW cell proliferative disorder; cancer; autoimmune; inflammatory disorder;
 KW allergy; anaemia; asthma; infection; reproductive disorder; infertility;
 KW ovulatory defect; neurological disorder; Alzheimer's disease;
 KW Parkinson's disease; gene therapy; ss.
 XX
 OS Homo sapiens.
 XX
 FH Key Location/Qualifiers
 FT CDS 1969..2439
 FT /tag=a
 FT /product="Carbohydrate-associated protein (CARBAP) -2"
 FT
 XX
 PN WO200229055-A2.
 XX
 XX 11-APR-2002.
 XX
 PF 01-OCT-2001; 2001MO-US30591.
 XX
 PF 02-OCT-2000; 2000US-237456P.
 XX
 XX (INCY-) INCYTE GENOMICS INC.
 PA
 PI Tang YT, Elliott VS, Yue H, Lal P, Malia NK;
 PI WPI; 2002-426114/45.
 DR P-PSDB; AAE14720.
 XX
 XX Novel human carbohydrate-associated proteins and genes, useful in the
 PT diagnosis, prevention and treatment of cell proliferative,
 PT autoimmune/inflammatory, reproductive and neurological disorders -
 XX
 XX Claim 5; Page 97-98; 98pp; English.
 XX
 CC The present sequence is human carbohydrate-associated protein (CARBAP) -2
 CC cDNA. The CARBAP and the polynucleotide encoding it are useful for
 CC diagnosing, treating and preventing cell proliferative disorders (e.g.
 CC arteriosclerosis, cirrhosis, hepatitis, cancer), autoimmune/inflammatory
 CC disorders (e.g. allergies, anaemia, asthma, osteoporosis, rheumatoid
 CC arthritis, atopic dermatitis, glomerulonephritis and irritable bowel
 CC syndrome, trauma, and bacterial, viral, parasitic, protozoal fungal or
 CC helminthic infections), reproductive disorders (e.g. infertility,
 CC ovulatory defects and fibrocystic breast disease), and neurological
 CC disorders (e.g. epilepsy, stroke, Alzheimer's disease, Huntington's
 CC disease, dementia, Parkinson's disease, multiple sclerosis, mental
 CC disorders including mood, anxiety and schizophrenic disorders, kuru,
 CC bacterial and viral meningitis, Pick's disease and amyotrophic lateral
 CC sclerosis). The polypeptide of the invention is also useful for screening
 CC agonist, antagonist, a compound that specifically binds to it or
 CC modulates its activity, for preparing polyclonal or monoclonal
 CC antibodies, in a number of drug screening techniques, to analyse the
 CC presence of a tissue or cell type and as element on a microarray.
 CC The polynucleotide is useful for creating knockin humanised animals or
 CC transgenic animals to model human diseases, in somatic or germline gene
 CC therapy, to generate a transcript image of a tissue or cell type, for
 CC detecting differences in the chromosomal location due to translocation,
 CC inversion, among normal, carrier or affected individuals, and as

CC hybridisation probes for mapping naturally occurring genomic sequences.
 XX
 SQ Sequence 3239 BP; 659 A; 1005 C; 814 G; 761 T; 0 other;

Query Match 2.1%; Score 29; DB 24; Length 3239;
 Best Local Similarity 100.0%; Pred. No. 0.071;
 Matches 29; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1369 AAAAAAAAAAAAAAAAAAGCGGTC 1397
 |||||
 Db 38 AAAAAAAAAAAAAAAAAAGCGGTC 10

Search completed: June 29, 2003, 05:30:21
 job time : 268 secs

831	23	1.6	646	4	US-09-227-357-101	Sequence 101, App	904	23	1.6	841	5	PCT-US95-14442A-110	Sequence 110, App
C 832	23	1.6	648	4	US-09-385-982-332	Sequence 332, App	905	23	1.6	842	2	US-08-924-759-1	Sequence 1, App1
C 833	23	1.6	650	4	US-09-328-111-179	Sequence 179, App	906	23	1.6	844	2	US-09-248-335-1	Sequence 1, App1
C 834	23	1.6	653	4	US-09-385-982-324	Sequence 324, App	907	23	1.6	855	6	5195441-40	Patent No. 5195441
C 835	23	1.6	654	5	PCT-US95-06406A-11	Sequence 11, App1	908	23	1.6	856	3	5223394-3	Patent No. 5223394
C 836	23	1.6	656	4	US-09-328-111-587	Sequence 587, App	909	23	1.6	857	3	US-08-460-040-1	Sequence 1, App1
C 837	23	1.6	657	4	US-09-385-982-91	Sequence 91, App1	910	23	1.6	860	1	US-07-847-010-18	Sequence 18, App1
C 838	23	1.6	658	4	US-09-328-111-816	Sequence 816, App	911	23	1.6	863	1	US-07-940-861-11	Sequence 11, App1
C 839	23	1.6	658	4	US-09-385-982-327	Sequence 327, App	912	23	1.6	863	1	US-08-459-512-11	Sequence 11, App1
C 840	23	1.6	660	1	US-08-555-678-41	Sequence 41, App1	913	23	1.6	863	2	US-08-459-657-11	Sequence 11, App1
C 841	23	1.6	661	4	US-09-328-111-69	Sequence 69, App1	914	23	1.6	863	2	US-08-460-132-11	Sequence 11, App1
C 842	23	1.6	669	4	US-09-328-111-342	Sequence 342, App	915	23	1.6	863	4	US-08-943-731-171	Sequence 171, App
C 843	23	1.6	669	4	US-09-328-111-782	Sequence 782, App	916	23	1.6	863	5	PCT-US92-02050-11	Sequence 11, App1
C 844	23	1.6	672	1	US-07-847-010-21	Sequence 21, App1	917	23	1.6	863	5	5195441-35	Patent No. 5195441
C 845	23	1.6	676	4	US-09-328-111-652	Sequence 652, App	918	23	1.6	863	6	5223394-5	Patent No. 5223394
C 846	23	1.6	683	1	US-07-972-481-1	Sequence 1, App1	919	23	1.6	870	5	PCT-US95-13658-3	Sequence 3, App1
C 847	23	1.6	689	4	US-09-149-476-25	Sequence 25, App1	920	23	1.6	878	4	US-09-300-6818-3	Sequence 3, App1
C 848	23	1.6	689	4	US-09-105-542A-14	Sequence 14, App1	921	23	1.6	882	4	US-09-068-140A-1	Sequence 1, App1
C 849	23	1.6	690	4	US-09-328-111-74	Sequence 74, App1	922	23	1.6	893	4	US-09-370-838-166	Sequence 166, App
C 850	23	1.6	692	4	US-09-328-111-285	Sequence 285, App1	923	23	1.6	900	3	US-09-248-335-61	Sequence 61, App1
C 851	23	1.6	704	4	US-09-475-316A-26	Sequence 26, App1	924	23	1.6	903	5	PCT-US95-06406A-21	Sequence 21, App1
C 852	23	1.6	704	4	US-08-896-164-49	Sequence 49, App1	925	23	1.6	911	2	US-08-924-759-9	Sequence 9, App1
C 853	23	1.6	706	4	US-09-191-136-14	Sequence 14, App1	926	23	1.6	911	3	US-09-248-335-9	Sequence 9, App1
C 854	23	1.6	717	4	US-08-988-416-904	Sequence 904, App1	927	23	1.6	923	3	US-08-906-769-144	Sequence 144, App
C 855	23	1.6	724	4	US-09-020-956-38	Sequence 38, App1	928	23	1.6	923	3	US-08-906-616-144	Sequence 144, App
C 856	23	1.6	724	4	US-09-030-607-38	Sequence 38, App1	929	23	1.6	923	3	US-08-639-075A-144	Sequence 144, App
C 857	23	1.6	724	4	US-09-605-785-38	Sequence 38, App1	930	23	1.6	923	4	US-09-004-731-66	Sequence 66, App1
C 858	23	1.6	724	4	US-09-439-313-38	Sequence 38, App1	931	23	1.6	923	4	US-09-012-431-144	Sequence 144, App
C 859	23	1.6	724	4	US-09-724-864-23	Sequence 23, App1	932	23	1.6	923	4	US-08-749-699-66	Sequence 66, App1
C 860	23	1.6	724	4	US-09-352-616A-38	Sequence 38, App1	933	23	1.6	923	4	US-09-012-692-144	Sequence 144, App
C 861	23	1.6	724	4	US-09-232-145A-38	Sequence 38, App1	934	23	1.6	923	4	US-08-906-613-144	Sequence 144, App
C 862	23	1.6	732	4	US-09-328-111-697	Sequence 697, App	935	23	1.6	923	4	US-09-336-643A-15	Sequence 15, App1
C 863	23	1.6	732	4	US-09-437-054A-15	Sequence 15, App1	936	23	1.6	923	4	US-09-004-729-66	Sequence 66, App1
C 864	23	1.6	740	2	US-08-713-000-8	Sequence 8, App1	937	23	1.6	936	2	US-08-557-309B-21	Sequence 21, App1
C 865	23	1.6	740	2	US-08-975-316-8	Sequence 8, App1	938	23	1.6	936	3	US-08-834-306-21	Sequence 21, App1
C 866	23	1.6	740	4	US-09-211-710-8	Sequence 8, App1	939	23	1.6	936	4	US-08-993-674A-21	Sequence 21, App1
C 867	23	1.6	740	4	US-09-615-192A-8	Sequence 8, App1	940	23	1.6	936	4	US-09-256-976-21	Sequence 21, App1
C 868	23	1.6	741	4	US-08-975-316-58	Sequence 58, App1	941	23	1.6	946	3	US-08-916-443A-15	Sequence 15, App1
C 869	23	1.6	741	4	US-09-615-192A-58	Sequence 58, App1	942	23	1.6	946	4	US-09-410-464-2	Sequence 2, App1
C 870	23	1.6	746	4	US-09-013-810-1	Sequence 1, App1	943	23	1.6	950	4	US-09-593-995-3	Sequence 3, App1
C 871	23	1.6	748	1	US-08-361-67B-3	Sequence 3, App1	944	23	1.6	958	2	US-08-757-046A-5	Sequence 5, App1
C 872	23	1.6	753	4	US-09-152-060-34	Sequence 34, App1	945	23	1.6	958	3	US-09-447-208-5	Sequence 5, App1
C 873	23	1.6	753	4	US-09-385-982-536	Sequence 536, App	946	23	1.6	958	4	US-09-135-988-5	Sequence 5, App1
C 874	23	1.6	760	2	US-08-276-452A-49	Sequence 49, App1	947	23	1.6	958	4	US-09-277-716-5	Sequence 5, App1
C 875	23	1.6	760	2	US-08-798-744-49	Sequence 49, App1	948	23	1.6	958	4	US-08-597-274A-5	Sequence 5, App1
C 876	23	1.6	767	1	US-08-184-604-1	Sequence 1, App1	949	23	1.6	958	4	US-08-808-909-5	Sequence 5, App1
C 877	23	1.6	772	4	US-09-575-602-1	Sequence 1, App1	950	23	1.6	958	4	US-09-609-161B-5	Sequence 5, App1
C 878	23	1.6	777	4	US-09-385-982-531	Sequence 531, App	951	23	1.6	961	6	5194596-16	Patent No. 5194596
C 879	23	1.6	777	4	US-09-149-476-181	Sequence 181, App	952	23	1.6	961	6	5219739-16	Patent No. 5219739
C 880	23	1.6	788	4	US-09-615-192A-96	Sequence 96, App1	953	23	1.6	966	1	US-08-514-014-7	Sequence 7, App1
C 881	23	1.6	788	4	5219739-8	Patent No. 5219739	954	23	1.6	966	2	US-08-833-823-7	Sequence 7, App1
C 882	23	1.6	790	6	5194596-8	Patent No. 5194596	955	23	1.6	975	4	US-09-535-008-39	Sequence 39, App1
C 883	23	1.6	794	4	US-09-385-982-537	Sequence 537, App	956	23	1.6	975	4	US-09-381-488-6	Sequence 6, App1
C 884	23	1.6	796	1	US-08-104-073-2	Sequence 2, App1	957	23	1.6	976	2	US-09-045-494-3	Sequence 3, App1
C 885	23	1.6	808	3	US-08-651-136C-15	Sequence 15, App1	958	23	1.6	976	2	US-08-504-459-9	Sequence 9, App1
C 886	23	1.6	808	3	US-09-229-911A-15	Sequence 15, App1	959	23	1.6	980	4	US-08-684-862-10	Sequence 10, App1
C 887	23	1.6	812	3	US-08-462-778-1	Sequence 1, App1	960	23	1.6	980	4	US-08-684-862-10	Sequence 10, App1
C 888	23	1.6	812	3	US-08-553-336A-1	Sequence 1, App1	961	23	1.6	981	3	US-08-924-747-25	Sequence 25, App1
C 889	23	1.6	813	4	US-09-310-838-29	Sequence 29, App1	962	23	1.6	991	4	US-09-247-373B-25	Sequence 25, App1
C 890	23	1.6	822	1	US-07-644-372-1	Sequence 1, App1	963	23	1.6	991	4	US-09-296-715-25	Sequence 25, App1
C 891	23	1.6	827	5	PCT-US95-06406A-20	Sequence 20, App1	964	23	1.6	996	2	US-08-624-650-2	Sequence 2, App1
C 892	23	1.6	840	3	US-08-924-759-13	Sequence 13, App1	965	23	1.6	1000	4	US-09-018-58A-33	Sequence 33, App1
C 893	23	1.6	840	3	US-09-348-335-13	Sequence 13, App1	966	23	1.6	1001	4	US-09-641-638-115	Sequence 115, App
C 894	23	1.6	841	3	US-08-906-769-110	Sequence 110, App	967	23	1.6	1001	4	US-09-641-638-115	Sequence 115, App
C 895	23	1.6	841	3	US-08-906-616-110	Sequence 110, App	968	23	1.6	1001	4	US-09-641-638-115	Sequence 115, App
C 896	23	1.6	841	3	US-08-817-795-110	Sequence 110, App	969	23	1.6	1001	4	US-09-641-638-115	Sequence 115, App
C 897	23	1.6	841	3	US-08-639-075A-110	Sequence 110, App	970	23	1.6	1003	2	US-08-887-997B-61	Sequence 61, App1
C 898	23	1.6	841	3	US-09-012-431-110	Sequence 110, App	971	23	1.6	1007	4	US-09-171-156A-61	Sequence 61, App1
C 899	23	1.6	841	4	US-09-032-215-36	Sequence 36, App1	972	23	1.6	1007	4	US-09-171-156A-61	Sequence 61, App1
C 900	23	1.6	841	4	US-09-032-215-36	Sequence 36, App1	973	23	1.6	1008	1	US-08-252-966B-15	Sequence 15, App1
C 901	23	1.6	841	4	US-09-012-692-110	Sequence 110, App	974	23	1.6	1008	4	US-09-091-097-38	Sequence 38, App1
C 902	23	1.6	841	4	US-08-906-613-110	Sequence 110, App	975	23	1.6	1009	6	5223394-8	Patent No. 5223394
C 903	23	1.6	841	4	US-08-906-613-110	Sequence 110, App	976	23	1.6	1009	6	5223394-8	Patent No. 5223394

977 23 1.6 1013 3 US-09-248-335-71. Sequence 71, Appl
978 23 1.6 1018 4 US-09-302-769-39 Sequence 39, Appl
979 23 1.6 1020 4 US-09-425-578-5 Sequence 5, Appl
980 23 1.6 1029 1 US-09-031-485-6 Sequence 6, Appl
981 23 1.6 1029 1 US-09-031-485-8 Sequence 8, Appl
982 23 1.6 1029 1 US-08-847-429A-6 Sequence 8, Appl
983 23 1.6 1029 1 US-08-847-429A-8 Sequence 8, Appl
984 23 1.6 1029 3 US-09-065-474-6 Sequence 6, Appl
985 23 1.6 1029 3 US-09-065-474-8 Sequence 8, Appl
986 23 1.6 1029 4 US-09-557-034-6 Sequence 8, Appl
987 23 1.6 1029 4 US-09-557-034-8 Sequence 8, Appl
988 23 1.6 1035 2 US-08-580-545B-3 Sequence 3, Appl
989 23 1.6 1035 4 US-09-262-653A-3 Sequence 3, Appl
990 23 1.6 1035 4 US-08-667-484A-1 Sequence 1, Appl
991 23 1.6 1037 2 US-08-824-405-1 Sequence 1, Appl
992 23 1.6 1046 4 US-09-068-140A-5 Sequence 5, Appl
993 23 1.6 1058 4 US-09-452-239-11 Sequence 11, Appl
994 23 1.6 1066 1 US-08-157-101A-4 Sequence 4, Appl
995 23 1.6 1074 3 US-09-248-335-67 Sequence 67, Appl
996 23 1.6 1075 4 US-08-400-006B-6 Sequence 6, Appl
997 23 1.6 1078 6 5223394-10 Patent No. 5223394
998 23 1.6 1082 4 US-09-247-155-181 Patent 181, App
999 23 1.6 1092 4 US-09-227-357-35 Sequence 35, App
1000 23 1.6 1100 3 US-09-248-335-53 Sequence 53, Appl

ALIGNMENTS

RESULT 1
US-08-473-981A-5
Sequence 5, Application US/08473981A
Patent No. 5629162
GENERAL INFORMATION:
APPLICANT: defougerolles, Antonin R
TITLE OF INVENTION: METHODS OF IDENTIFYING AGENTS WHICH MODULATE
TITLE OF INVENTION: ICAM-3 BINDING TO LFA-1 (AS AMENDED)
NUMBER OF SEQUENCES: 14
CORRESPONDENCE ADDRESS:
ADDRESSEE: STERNE, KESSLER, GOLDSTEIN & FOX, P.L.L.C.
STREET: 1100 NEW YORK AVENUE, N.W. SUITE 600
CITY: WASHINGTON
STATE: D. C.
COUNTRY: USA
ZIP: 20005
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/473,981A
FILING DATE: 07-JUN-1995
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: MILTONIG, ROBERT C
REGISTRATION NUMBER: 34,395
REFERENCE/DOCKET NUMBER: 1011.0560004
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202) 371-2600
TELEFAX: (202) 371-2540
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 1817 base pairs
TYPE: nucleic acid
STRANDEDNESS: both
TOPOLOGY: both
MOLECULE TYPE: DNA (genomic)
FEATURE:
NAME/KEY: CDS
LOCATION: 9..1649
US-08-473-981A-5

Query Match 1.9%; Score 27; DB 1; Length 1817;
Best Local Similarity 100.0%; Pred. No. 0.017;
Matches 27; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1364 TCCCTAAAAAAAAAAAAAAAAAAAAA 1390
|||||
Db 1727 TCCCTAAAAAAAAAAAAAAAAAAAAA 1753

RESULT 2
US-08-474-087-5
Sequence 5, Application US/08474087
Patent No. 5891841
GENERAL INFORMATION:
APPLICANT: de Fougerolles, Antonin R
TITLE OF INVENTION: METHODS OF USING INTERCELLULAR ADHESION MOLECULE-
NUMBER OF SEQUENCES: 14
CORRESPONDENCE ADDRESS:
ADDRESSEE: STERNE, KESSLER, GOLDSTEIN & FOX, P.L.L.C.
STREET: 1100 NEW YORK AVENUE, N.W. SUITE 600
CITY: WASHINGTON
STATE: D. C.
COUNTRY: USA
ZIP: 20005
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/474,087
FILING DATE: 07-JUN-1995
CLASSIFICATION: 514
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/038,990
FILING DATE: 23-DEC-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/712,879
FILING DATE: 11-JUN-1991
ATTORNEY/AGENT INFORMATION:
NAME: MILTONIG, ROBERT C
REGISTRATION NUMBER: 34,395
REFERENCE/DOCKET NUMBER: 1011.0560003
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202) 371-2600
TELEFAX: (202) 371-2540
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 1817 base pairs
TYPE: nucleic acid
STRANDEDNESS: both
TOPOLOGY: both
MOLECULE TYPE: DNA (genomic)
FEATURE:
NAME/KEY: CDS
LOCATION: 9..1649
US-08-474-087-5

Query Match 1.9%; Score 27; DB 2; Length 1817;
Best Local Similarity 100.0%; Pred. No. 0.017;
Matches 27; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1364 TCCCTAAAAAAAAAAAAAAAAAAAAA 1390
|||||
Db 1727 TCCCTAAAAAAAAAAAAAAAAAAAAA 1753

RESULT 3
US-09-130-491-7
Sequence 7, Application US/09130491
Patent No. 6416974

GENERAL INFORMATION:
APPLICANT: Holtzman, Douglas A.
APPLICANT: Goodheart, Andrew D.
TITLE OF INVENTION: TANGO-71, TANGO-73, TANGO-74, TANGO-76, AND TANGO-83
FILE REFERENCE: 09404/041001.
CURRENT APPLICATION NUMBER: US/09/130,491
EARLIER FILING DATE: 1998-08-07
EARLIER APPLICATION NUMBER: US 60/058,108
EARLIER FILING DATE: 1997-09-05
EARLIER APPLICATION NUMBER: US 60/054,961
EARLIER FILING DATE: 1997-08-06
NUMBER OF SEQ ID NOS: 16
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 7
LENGTH: 2114
TYPE: DNA
ORGANISM: Rattus rattus
FEATURE:
NAME/KEY: CDS
LOCATION: (3)...(1445)
US-09-130-491-7

Query Match 1.9%; Score 27; DB 4; Length 2114;
Best Local Similarity 100.0%; Pred. No. 0.017;
Matches 27; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1369 AAAAAAAAAAAAAAAAAAGCGG 1395
DB 2084 AAAAAAAAAAAAAAAAAAGCGG 2110

RESULT 4
US-09-798-096-10/C
Sequence 10, Application US/09798096
Patent No. 6393378
GENERAL INFORMATION:
APPLICANT: Donna T. Ward
APPLICANT: Andrew T. Walt
TITLE OF INVENTION: ANTISENSE MODULATION OF RECOL2 EXPRESSION
FILE REFERENCE: RTS-0207
CURRENT APPLICATION NUMBER: US/09/798,096
CURRENT FILING DATE: 2001-03-01
NUMBER OF SEQ ID NOS: 89
SEQ ID NO 10
LENGTH: 99500
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
US-09-798-096-10

Query Match 1.9%; Score 27; DB 4; Length 99500;
Best Local Similarity 100.0%; Pred. No. 0.014;
Matches 27; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1367 CTAATAAAAAAAAAAAAAAAAAAGGC 1393
DB 57302 CTAATAAAAAAAAAAAAAAAAAAGGC 57276

RESULT 5
US-09-122-400B-11
Sequence 11, Application US/09122400B
Patent No. 6245974
GENERAL INFORMATION:
APPLICANT: Michalowski, Susan
APPLICANT: Spiker, Steven
TITLE OF INVENTION: MATRIX ATTACHMENT REGIONS
FILE REFERENCE: Michalowski and Spiker
CURRENT APPLICATION NUMBER: US/09/122,400B
CURRENT FILING DATE: 1998-07-24
PRIOR APPLICATION NUMBER: 60/066,118
PRIOR FILING DATE: 1997-08-06
NUMBER OF SEQ ID NOS: 22

SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 11
LENGTH: 899
TYPE: DNA
ORGANISM: Nicotiana tabacum
US-09-122-400B-11

Query Match 1.9%; Score 26; DB 4; Length 899;
Best Local Similarity 100.0%; Pred. No. 0.047;
Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1365 CCTTAAAAAAAAAAAAAAAAAAAAA 1390
DB 804 CCTTAAAAAAAAAAAAAAAAAAAAA 829

RESULT 6
US-09-257-179-28
Sequence 28, Application US/09257179
Patent No. 6410709
GENERAL INFORMATION:
APPLICANT: Ruben et al.
TITLE OF INVENTION: 29 Human Secreted Proteins
FILE REFERENCE: P2015P1
CURRENT APPLICATION NUMBER: US/09/257,179
CURRENT FILING DATE: 1999-02-25
EARLIER APPLICATION NUMBER: PCT/US98/17709
EARLIER FILING DATE: 1998-08-27
EARLIER APPLICATION NUMBER: 60/056,270
EARLIER FILING DATE: 1997-08-29
EARLIER APPLICATION NUMBER: 60/056,271
EARLIER FILING DATE: 1997-08-29
EARLIER APPLICATION NUMBER: 60/056,247
EARLIER FILING DATE: 1997-08-29
EARLIER APPLICATION NUMBER: 60/056,073
EARLIER FILING DATE: 1997-08-29
NUMBER OF SEQ ID NOS: 128
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 28
LENGTH: 1327
TYPE: DNA
ORGANISM: Homo sapiens
US-09-257-179-28

Query Match 1.9%; Score 26; DB 4; Length 1327;
Best Local Similarity 100.0%; Pred. No. 0.046;
Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1370 AAAAAAAAAAAAAAAAAAGCGG 1395
DB 1300 AAAAAAAAAAAAAAAAAAGCGG 1325

RESULT 7
US-08-836-567-3
Sequence 3, Application US/08836567
Patent No. 6130367
GENERAL INFORMATION:
APPLICANT: Kossmann, Jens
APPLICANT: Springer, Franziska
APPLICANT: Abel, Gernot
TITLE OF INVENTION: DNA MOLECULES THAT CODE FOR ENZYMES INVOLVED IN STARCH SYNTHESIS VECTORS BACTERIA TRANSGENIC
TITLE OF INVENTION: PLANT CELLS AND PLANTS CONTAINING SAID MOLECULES
NUMBER OF SEQUENCES: 17
CORRESPONDENCE ADDRESS:
ADDRESSEE: FISH & NEAVE
STREET: 1251 Avenue of the Americas
CITY: New York
STATE: New York
COUNTRY: USA
ZIP: 10020
COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy/disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/836,567
FILING DATE: 24-JUL-1997
CLASSIFICATION: 800
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/EP95/04415
FILING DATE: 09-NOV-1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER: DE P 44 41 408.0
FILING DATE: 10-NOV-1994
ATTORNEY/AGENT INFORMATION:
NAME: Haley Jr., James F.
REGISTRATION NUMBER: 27,794
REFERENCE/DOCKET NUMBER: Agrevo-4
TELECOMMUNICATION INFORMATION:
TELEPHONE: 212-596-9000
TELEFAX: 212-596-9090
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 1758 base pairs
TYPE: nucleotide
STRANDEDNESS: unknown
TOPOLOGY: linear
MOLECULE TYPE: cDNA to mRNA
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: Solanum tuberosum
STRAIN: cv. Bercolina
TISSUE TYPE: tuber tissue
IMMEDIATE SOURCE:
LIBRARY: cDNA-library in pluescriptskit+
FEATURE:
NAME/KEY: CDS
LOCATION: 1..1377
OTHER INFORMATION: /function="Polymerization of starch"
OTHER INFORMATION: /product="Starch synthase"
US-08-836-567-3

Query Match 1.9%; Score 26; DB 3; Length 1758;
Best Local Similarity 100.0%; Pred. No. 0.045;
Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1365 CCTAATAAAAAAAAAAAAAAAAAA 1390
DB 1719 CCTAATAAAAAAAAAAAAAAAAAA 1744

RESULT 8
US-08-287-001A-1
Sequence 1, Application US/08287001A
Patent No. 5622861
GENERAL INFORMATION:
APPLICANT: KAPLAN, GERARDO
APPLICANT: FEINSTONE, STEPHEN M.
TITLE OF INVENTION: HEPATITIS A VIRUS RECEPTOR AND METHODS
NUMBER OF SEQUENCES: 5
CORRESPONDENCE ADDRESS:
ADDRESSEE: NEEDLE & ROSENBERG, P.C.
STREET: Suite 1200, The Candler Bldg, 127 Peachtree
STREET: Street
CITY: Atlanta
STATE: Georgia
COUNTRY: USA
ZIP: 30303
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/287,001A
FILING DATE:
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Spratt, Gwedolyn D.
REGISTRATION NUMBER: 36,016
REFERENCE/DOCKET NUMBER: 1414.621
TELECOMMUNICATION INFORMATION:
TELEPHONE: 404/688-0770
TELEFAX: 404/688-9880
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 2093 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA
FEATURE:
NAME/KEY: CDS
LOCATION: 196..1551
US-08-287-001A-1

Query Match 1.9%; Score 26; DB 1; Length 2093;
Best Local Similarity 100.0%; Pred. No. 0.045;
Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1365 CCTAATAAAAAAAAAAAAAAAAAA 1390
DB 2068 CCTAATAAAAAAAAAAAAAAAAAA 2093

RESULT 9
PCT-US95-09941-1
Sequence 1, Application PC/TUS9509941
GENERAL INFORMATION:
APPLICANT:
TITLE OF INVENTION: HEPATITIS A VIRUS RECEPTOR AND METHODS
NUMBER OF SEQUENCES: 5
CORRESPONDENCE ADDRESS:
ADDRESSEE: NEEDLE & ROSENBERG, P.C.
STREET: Suite 1200, The Candler Bldg, 127 Peachtree
STREET: Street
CITY: Atlanta
STATE: Georgia
COUNTRY: USA
ZIP: 30303
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US95/09941
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/287,001
FILING DATE: 5 AUG 1994
ATTORNEY/AGENT INFORMATION:
NAME: Spratt, Gwedolyn D.
REGISTRATION NUMBER: 36,016
REFERENCE/DOCKET NUMBER: 1414.621
TELECOMMUNICATION INFORMATION:
TELEPHONE: 404/688-0770
TELEFAX: 404/688-9880
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 2093 base pairs

TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
FEATURE:
NAME/KEY: CDS
LOCATION: 196..1551
PCT-US95-09941-1

Query Match 1.9%; Score 26; DB 5; Length 2093;
Best Local Similarity 100.0%; Pred. No. 0.04%;
Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1365 CCTAAAAAAAAAAAAAAAAAAAAA 1390
DB 2068 CCTAAAAAAAAAAAAAAAAAAAAA 2093

RESULT 10
US-08-969-630-3
Sequence 3, Application US/08969630A
Patent No. 5981248
GENERAL INFORMATION:
APPLICANT: Xu, Hua
TITLE OF INVENTION: MAMMALIAN CELL DEATH PREVENTING KINASE, DPK
FILE REFERENCE: A463
CURRENT APPLICATION NUMBER: US/08/969,630A
CURRENT FILING DATE: 1997-11-13
NUMBER OF SEQ ID NOS: 7
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 3
LENGTH: 3891
TYPE: DNA
ORGANISM: mouse
FEATURE:
NAME/KEY: CDS
LOCATION: (63)..(1565)
US-08-969-630-3

Query Match 1.9%; Score 26; DB 2; Length 3891;
Best Local Similarity 100.0%; Pred. No. 0.04%;
Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1365 CCTAAAAAAAAAAAAAAAAAAAAA 1390
DB 3861 CCTAAAAAAAAAAAAAAAAAAAAA 3886

RESULT 11
US-08-469-260A-11/c
Sequence 11, Application US/08469260A
Patent No. 6451578
GENERAL INFORMATION:
APPLICANT: JOHN N. SIMONS
APPLICANT: TAMI J. PILOT-MATIAS
APPLICANT: GEORGE J. DAWSON
APPLICANT: GEORGE G. SCHLAUDER
APPLICANT: SURESH M. DESAI
APPLICANT: THOMAS P. LEARY
APPLICANT: ANTHONY SCOTT MUEHROFF
APPLICANT: JAMES C. ERKER
APPLICANT: SHERI L. BUIJK
APPLICANT: ISA K. MUSHAWAR
TITLE OF INVENTION: NON-A, NON-B, NON-C, NON-D, NON-E HEPATITIS
TITLE OF INVENTION: REAGENTS AND METHODS FOR THEIR USE
NUMBER OF SEQUENCES: 716
CORRESPONDENCE ADDRESS:
ADDRESSEE: ABBOTT LABORATORIES D377/AP6D
STREET: 100 ABBOTT PARK ROAD
CITY: ABBOTT PARK
STATE: IL
COUNTRY: USA
ZIP: 60064-3500

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/469,260A
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/08/424,550
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: POREMBSKI, PRISCILLA E.
REGISTRATION NUMBER: 33,207
REFERENCE/DOCKET NUMBER: 5527.PC.01
TELECOMMUNICATION INFORMATION:
TELEPHONE: 708-937-6365
TELEFAX: 708-938-2623
INFORMATION FOR SEQ ID NO: 11:
SEQUENCE CHARACTERISTICS:
LENGTH: 8912 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
US-08-469-260A-11

Query Match 1.9%; Score 26; DB 4; Length 8912;
Best Local Similarity 100.0%; Pred. No. 0.04%;
Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1365 CCTAAAAAAAAAAAAAAAAAAAAA 1390
DB 8911 CCTAAAAAAAAAAAAAAAAAAAAA 8886

RESULT 12
US-08-639-857-32/c
Sequence 32, Application US/08639857
Patent No. 5955318
GENERAL INFORMATION:
APPLICANT: Simons, J. N.
APPLICANT: Desai, S. M.
APPLICANT: Mushawar, I. K.
TITLE OF INVENTION: REAGENTS AND METHODS USEFUL FOR CONTROLLING THE
TITLE OF INVENTION: TRANSLATION OF HEPATITIS GB PROTEINS
NUMBER OF SEQUENCES: 32
CORRESPONDENCE ADDRESS:
ADDRESSEE: Abbott Laboratories
STREET: 100 Abbott Park Rd
CITY: Abbott Park
STATE: IL
COUNTRY: USA
ZIP: 60064
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/639,857
FILING DATE:
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Porembski, Priscilla E.
REGISTRATION NUMBER: 33,207
REFERENCE/DOCKET NUMBER: 5793.US.P1
TELECOMMUNICATION INFORMATION:
TELEPHONE: 708-937-0378
TELEFAX: 708-938-2623
INFORMATION FOR SEQ ID NO: 32:
SEQUENCE CHARACTERISTICS:

LENGTH: 9143 base pairs
 TYPE: nucleic acid
 STRANDEDNESS: double
 TOPOLOGY: linear
 MOLECULE TYPE: DNA (genomic)
 US-08-639-857-32

Query Match 1.9%; Score 26; DB 2; Length 9143;
 Best Local Similarity 100.0%; Pred. No. 0.041;
 Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1365 CCTAAAAAAAAAAAAAAAAAAAAA 1390
 DB 9098 CCTAAAAAAAAAAAAAAAAAAAAA 9073

RESULT 13
 US-08-469-260A-390/C
 Sequence 390, Application US/08469260A
 Patent No. 6451578

GENERAL INFORMATION:
 APPLICANT: JOHN N. SIMONS
 APPLICANT: TAMI J. PILOT-MATIAS
 APPLICANT: GEORGE J. DAMSON
 APPLICANT: GEORGE G. SCHLAUDER
 APPLICANT: SURESH M. DESAI
 APPLICANT: THOMAS P. LEARY
 APPLICANT: ANTHONY SCOTT MUEHROFF
 APPLICANT: JAMES C. ERKER
 APPLICANT: SHERI L. BUTIK
 APPLICANT: ISA K. MUSHAMMAR
 TITLE OF INVENTION: NON-A, NON-B, NON-C, NON-D, NON-E HEPATITIS
 TITLE OF INVENTION: REAGENTS AND METHODS FOR THEIR USE
 NUMBER OF SEQUENCES: 716
 CORRESPONDENCE ADDRESSES:
 ADDRESSEE: ABBOTT LABORATORIES D377/AP6D
 STREET: 100 ABBOTT PARK ROAD
 CITY: ABBOTT PARK
 STATE: IL
 COUNTRY: USA
 ZIP: 60064-3500
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: Patent Release #1.0, Version #1.25
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/469,260A
 FILING DATE:
 CLASSIFICATION:
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US/08/424,550
 FILING DATE:
 ATTORNEY/AGENT INFORMATION:
 NAME: FOREMSKI, PRISCILLA E.
 REGISTRATION NUMBER: 33,207
 REFERENCE/DOCKET NUMBER: 5527.PC.01
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: 708-937-6365
 TELEFAX: 708-938-2623
 INFORMATION FOR SEQ ID NO: 390:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 9143 base pairs
 TYPE: nucleic acid
 STRANDEDNESS: double
 TOPOLOGY: linear
 MOLECULE TYPE: DNA (genomic)
 US-08-469-260A-390.

Query Match 1.9%; Score 26; DB 4; Length 9143;
 Best Local Similarity 100.0%; Pred. No. 0.041;
 Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1365 CCTAAAAAAAAAAAAAAAAAAAAA 1390
 DB 9098 CCTAAAAAAAAAAAAAAAAAAAAA 9073

RESULT 14
 US-08-469-260A-393/C
 Sequence 393, Application US/08469260A
 Patent No. 6451578

GENERAL INFORMATION:
 APPLICANT: JOHN N. SIMONS
 APPLICANT: TAMI J. PILOT-MATIAS
 APPLICANT: GEORGE J. DAMSON
 APPLICANT: GEORGE G. SCHLAUDER
 APPLICANT: SURESH M. DESAI
 APPLICANT: THOMAS P. LEARY
 APPLICANT: ANTHONY SCOTT MUEHROFF
 APPLICANT: JAMES C. ERKER
 APPLICANT: SHERI L. BUTIK
 APPLICANT: ISA K. MUSHAMMAR
 TITLE OF INVENTION: NON-A, NON-B, NON-C, NON-D, NON-E HEPATITIS
 TITLE OF INVENTION: REAGENTS AND METHODS FOR THEIR USE
 NUMBER OF SEQUENCES: 716
 CORRESPONDENCE ADDRESSES:
 ADDRESSEE: ABBOTT LABORATORIES D377/AP6D
 STREET: 100 ABBOTT PARK ROAD
 CITY: ABBOTT PARK
 STATE: IL
 COUNTRY: USA
 ZIP: 60064-3500
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: Patent Release #1.0, Version #1.25
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/469,260A
 FILING DATE:
 CLASSIFICATION:
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US/08/424,550
 FILING DATE:
 ATTORNEY/AGENT INFORMATION:
 NAME: FOREMSKI, PRISCILLA E.
 REGISTRATION NUMBER: 33,207
 REFERENCE/DOCKET NUMBER: 5527.PC.01
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: 708-937-6365
 TELEFAX: 708-938-2623
 INFORMATION FOR SEQ ID NO: 393:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 9143 base pairs
 TYPE: nucleic acid
 STRANDEDNESS: double
 TOPOLOGY: linear
 MOLECULE TYPE: DNA (genomic)
 FEATURE:
 NAME/KEY: 5'UTR
 LOCATION: 1..445
 FEATURE:
 NAME/KEY: CDS
 LOCATION: 446..9037
 FEATURE:
 NAME/KEY: 3'UTR
 LOCATION: 9038..9143
 US-08-469-260A-393

Query Match 1.9%; Score 26; DB 4; Length 9143;
 Best Local Similarity 100.0%; Pred. No. 0.041;
 Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

DB 9098 CCTAAAAAAAAAAAAAAAAAAAAA 9073

RESULT 15

US-09-821-736-3
Sequence 3, Application US/09821736

Patent No. 6326182

GENERAL INFORMATION:

APPLICANT: WEBSTER, Marion et al

TITLE OF INVENTION: ISOLATED HUMAN LIPASE PROTEINS, NUCLEIC

TITLE OF INVENTION: ACID MOLECULES ENCODING HUMAN LIPASE PROTEINS, AND USES

TITLE OF INVENTION: THEREOF

FILE REFERENCE: CL001216

CURRENT APPLICATION NUMBER: US/09/821,736

CURRENT FILING DATE: 2001-03-30

NUMBER OF SEQ ID NOS: 5

SOFTWARE: FastSeq for Windows Version 4.0

SEQ ID NO 3

LENGTH: 14753

TYPE: DNA

ORGANISM: Human

FEATURE:

NAME/KEY: misc_feature

LOCATION: (1)..(14753)

OTHER INFORMATION: n = A,T,C or G

US-09-821-736-3

Query Match

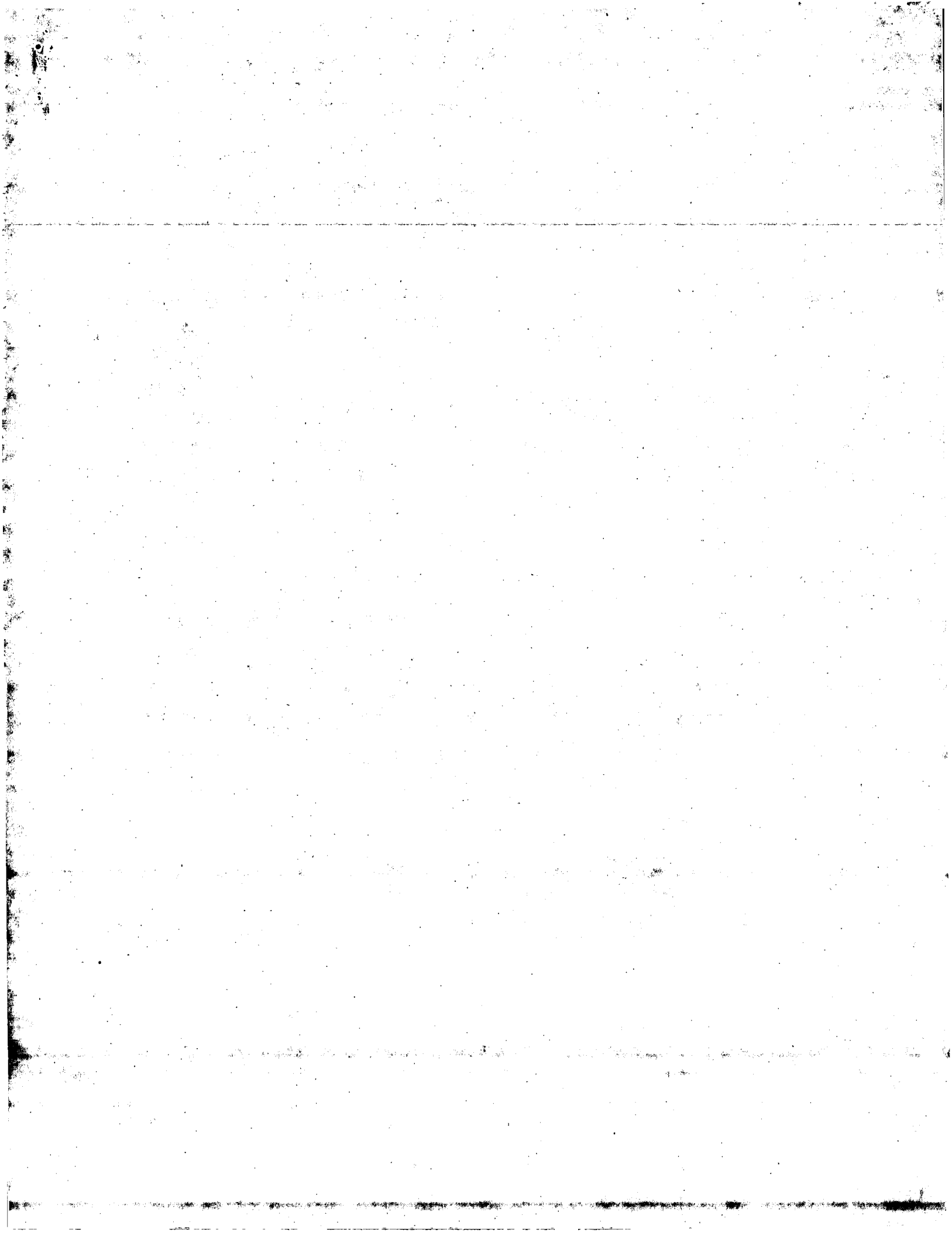
Best Local Similarity 1.9%; Score 26; DB 4; Length 14753;

Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1366 CCTAAAAAAAAAAAAAAAAAAAAAAG 1391

DB 2131 CCTAAAAAAAAAAAAAAAAAAAAAAG 2156

Search completed: June 29, 2003, 06:37:45
Job time : 75 secs



Mon Jun 30 09:37:15 2003

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 29, 2003, 05:25:43 ; Search time 166 Seconds
(without alignments)
12492.616 Million cell updates/sec

Title: US-09-989-919-15

Perfect score: 1397
Sequence: 1 ggtgtgtacgtctacgga.....aaaaaaaaaagcggtc 1397Scoring table: OLIGO_NUC
Gapop_60.0, Gapext 60.0

Searched: 1055720 seqs, 742224136 residues

Word size: 0

Total number of hits satisfying chosen parameters: 2111440

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 1000 summaries

Database:

```
1: /cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq:*
2: /cgn2_6/ptodata/1/pubpna/PCT_NEW_PUB.seq:*
3: /cgn2_6/ptodata/1/pubpna/US06_NEW_PUB.seq:*
4: /cgn2_6/ptodata/1/pubpna/US06_PUBCOMB.seq:*
5: /cgn2_6/ptodata/1/pubpna/US07_NEW_PUB.seq:*
6: /cgn2_6/ptodata/1/pubpna/PCTUS_PUBCOMB.seq:*
7: /cgn2_6/ptodata/1/pubpna/US08_NEW_PUB.seq:*
8: /cgn2_6/ptodata/1/pubpna/US08_PUBCOMB.seq:*
9: /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq:*
10: /cgn2_6/ptodata/1/pubpna/US05_PUBCOMB.seq:*
11: /cgn2_6/ptodata/1/pubpna/US10_NEW_PUB.seq:*
12: /cgn2_6/ptodata/1/pubpna/US10_PUBCOMB.seq:*
13: /cgn2_6/ptodata/1/pubpna/US60_NEW_PUB.seq:*
14: /cgn2_6/ptodata/1/pubpna/US60_PUBCOMB.seq:*
```

pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1397	100.0	470	9	US-09-989-919-15
2	442	31.6	437	9	US-09-989-919-14
3	322	23.0	427	10	US-09-989-919-14
4	97	6.9	493	9	US-09-918-995-32213
5	30	2.1	386	10	US-09-770-791-70
6	30	2.1	421	10	US-09-867-701-3641
7	30	2.1	777	10	US-09-770-445-884
8	30	2.1	2457	9	US-10-176-459-569
9	30	2.1	2457	9	US-10-176-558-569
10	30	2.1	2457	9	US-10-175-737-569
11	30	2.1	2457	9	US-10-173-706-569
12	30	2.1	2457	9	US-10-175-738-569
13	30	2.1	2457	9	US-10-175-752-569
14	30	2.1	2457	9	US-10-176-482-569
15	30	2.1	2457	9	US-10-176-757-569
16	30	2.1	2457	9	US-10-176-913-569
17	30	2.1	2457	9	US-10-180-552-569
18	30	2.1	2457	9	US-10-180-557-569
19	30	2.1	2457	9	US-10-173-700-569

20	30	2.1	2457	9	US-10-174-572-569	Sequence 569, App
21	30	2.1	2457	9	US-10-174-579-569	Sequence 569, App
22	30	2.1	2457	9	US-10-174-582-569	Sequence 569, App
23	30	2.1	2457	9	US-10-174-588-569	Sequence 569, App
24	30	2.1	2457	9	US-10-175-729-569	Sequence 569, App
25	30	2.1	2457	9	US-10-175-740-569	Sequence 569, App
26	30	2.1	2457	9	US-10-175-743-569	Sequence 569, App
27	30	2.1	2457	9	US-10-176-489-569	Sequence 569, App
28	30	2.1	2457	9	US-10-176-492-569	Sequence 569, App
29	30	2.1	2457	9	US-10-176-747-569	Sequence 569, App
30	30	2.1	2457	9	US-10-176-750-569	Sequence 569, App
31	30	2.1	2457	9	US-10-176-985-569	Sequence 569, App
32	30	2.1	2457	9	US-10-176-987-569	Sequence 569, App
33	30	2.1	2457	9	US-10-176-991-569	Sequence 569, App
34	30	2.1	2457	9	US-10-176-992-569	Sequence 569, App
35	30	2.1	2457	9	US-10-176-993-569	Sequence 569, App
36	30	2.1	2457	9	US-10-184-658-569	Sequence 569, App
37	30	2.1	2457	9	US-10-173-695-569	Sequence 569, App
38	30	2.1	2457	9	US-10-173-705-569	Sequence 569, App
39	30	2.1	2457	9	US-10-173-705-569	Sequence 569, App
40	30	2.1	2457	9	US-10-174-576-569	Sequence 569, App
41	30	2.1	2457	9	US-10-174-585-569	Sequence 569, App
42	30	2.1	2457	9	US-10-174-586-569	Sequence 569, App
43	30	2.1	2457	9	US-10-175-747-569	Sequence 569, App
44	30	2.1	2457	9	US-10-176-481-569	Sequence 569, App
45	30	2.1	2457	9	US-10-176-485-569	Sequence 569, App
46	30	2.1	2457	9	US-10-176-487-569	Sequence 569, App
47	30	2.1	2457	9	US-10-176-493-569	Sequence 569, App
48	30	2.1	2457	9	US-10-176-736-569	Sequence 569, App
49	30	2.1	2457	9	US-10-176-911-569	Sequence 569, App
50	30	2.1	2457	9	US-10-176-919-569	Sequence 569, App
51	30	2.1	2457	9	US-10-176-925-569	Sequence 569, App
52	30	2.1	2457	9	US-10-176-978-569	Sequence 569, App
53	30	2.1	2457	9	US-10-179-510-569	Sequence 569, App
54	30	2.1	2457	9	US-10-180-543-569	Sequence 569, App
55	30	2.1	2457	9	US-10-180-544-569	Sequence 569, App
56	30	2.1	2457	9	US-10-180-546-569	Sequence 569, App
57	30	2.1	2457	9	US-10-180-547-569	Sequence 569, App
58	30	2.1	2457	9	US-10-180-549-569	Sequence 569, App
59	30	2.1	2457	9	US-10-180-555-569	Sequence 569, App
60	30	2.1	2457	9	US-10-180-558-569	Sequence 569, App
61	30	2.1	2457	9	US-10-181-000-569	Sequence 569, App
62	30	2.1	2457	9	US-10-183-010-569	Sequence 569, App
63	30	2.1	2457	9	US-10-183-012-569	Sequence 569, App
64	30	2.1	2457	9	US-10-184-614-569	Sequence 569, App
65	30	2.1	2457	9	US-10-184-623-569	Sequence 569, App
66	30	2.1	2457	9	US-10-184-635-569	Sequence 569, App
67	30	2.1	2457	9	US-10-184-637-569	Sequence 569, App
68	30	2.1	2457	9	US-10-184-646-569	Sequence 569, App
69	30	2.1	2457	9	US-10-184-647-569	Sequence 569, App
70	30	2.1	2457	9	US-10-184-653-569	Sequence 569, App
71	30	2.1	2457	9	US-10-187-594-569	Sequence 569, App
72	30	2.1	2457	9	US-10-187-596-569	Sequence 569, App
73	30	2.1	2457	9	US-10-187-745-569	Sequence 569, App
74	30	2.1	2457	9	US-10-187-745-569	Sequence 569, App
75	30	2.1	2457	9	US-10-187-886-569	Sequence 569, App
76	30	2.1	2457	9	US-10-199-464-569	Sequence 569, App
77	30	2.1	2457	9	US-10-176-751-569	Sequence 569, App
78	30	2.1	2457	9	US-10-176-760-569	Sequence 569, App
79	30	2.1	2457	9	US-10-176-990-569	Sequence 569, App
80	30	2.1	2457	9	US-10-180-541-569	Sequence 569, App
81	30	2.1	2457	9	US-10-180-542-569	Sequence 569, App
82	30	2.1	2457	9	US-10-180-544-569	Sequence 569, App
83	30	2.1	2457	9	US-10-180-551-569	Sequence 569, App
84	30	2.1	2457	9	US-10-180-998-569	Sequence 569, App
85	30	2.1	2457	9	US-10-180-999-569	Sequence 569, App
86	30	2.1	2457	9	US-10-183-013-569	Sequence 569, App
87	30	2.1	2457	9	US-10-184-612-569	Sequence 569, App
88	30	2.1	2457	9	US-10-184-615-569	Sequence 569, App
89	30	2.1	2457	9	US-10-184-617-569	Sequence 569, App
90	30	2.1	2457	9	US-10-184-621-569	Sequence 569, App
91	30	2.1	2457	9	US-10-184-628-569	Sequence 569, App
92	30	2.1	2457	9	US-10-184-629-569	Sequence 569, App

823	25	1.8	1656	9	US-10-140-474-239	Sequence 239, App
824	25	1.8	1656	9	US-10-142-431-239	Sequence 239, App
825	25	1.8	1656	9	US-10-143-114-239	Sequence 239, App
826	25	1.8	1656	9	US-10-140-002-239	Sequence 239, App
827	25	1.8	1656	9	US-10-142-419-239	Sequence 239, App
828	25	1.8	1656	9	US-10-123-262-239	Sequence 239, App
829	25	1.8	1656	9	US-10-142-423-239	Sequence 239, App
830	25	1.8	1656	9	US-10-121-050-239	Sequence 239, App
831	25	1.8	1656	9	US-10-141-755-239	Sequence 239, App
832	25	1.8	1656	9	US-10-143-032-239	Sequence 239, App
833	25	1.8	1656	9	US-10-123-108-239	Sequence 239, App
834	25	1.8	1656	9	US-10-123-236-239	Sequence 239, App
835	25	1.8	1656	9	US-10-123-261-239	Sequence 239, App
836	25	1.8	1656	9	US-10-140-921-239	Sequence 239, App
837	25	1.8	1656	9	US-10-140-928-239	Sequence 239, App
838	25	1.8	1656	9	US-10-121-045-239	Sequence 239, App
839	25	1.8	1656	9	US-10-123-292-239	Sequence 239, App
840	25	1.8	1656	9	US-10-123-503-239	Sequence 239, App
841	25	1.8	1656	9	US-10-124-819-239	Sequence 239, App
842	25	1.8	1656	9	US-10-124-822-239	Sequence 239, App
843	25	1.8	1656	9	US-10-140-925-239	Sequence 239, App
844	25	1.8	1656	9	US-10-160-498-239	Sequence 239, App
845	25	1.8	1656	9	US-10-121-041-239	Sequence 239, App
846	25	1.8	1656	9	US-10-121-043-239	Sequence 239, App
847	25	1.8	1656	9	US-10-121-047-239	Sequence 239, App
848	25	1.8	1656	9	US-10-123-215-239	Sequence 239, App
849	25	1.8	1656	9	US-10-123-902-239	Sequence 239, App
850	25	1.8	1656	9	US-10-123-908-239	Sequence 239, App
851	25	1.8	1656	9	US-10-123-909-239	Sequence 239, App
852	25	1.8	1656	9	US-10-123-910-239	Sequence 239, App
853	25	1.8	1656	9	US-10-124-813-239	Sequence 239, App
854	25	1.8	1656	9	US-10-124-817-239	Sequence 239, App
855	25	1.8	1656	9	US-10-124-824-239	Sequence 239, App
856	25	1.8	1656	9	US-10-125-922-239	Sequence 239, App
857	25	1.8	1656	9	US-10-125-924-239	Sequence 239, App
858	25	1.8	1656	9	US-10-127-825A-239	Sequence 239, App
859	25	1.8	1656	9	US-10-127-829A-239	Sequence 239, App
860	25	1.8	1656	9	US-10-127-835A-239	Sequence 239, App
861	25	1.8	1656	9	US-10-127-839A-239	Sequence 239, App
862	25	1.8	1656	9	US-10-127-901A-239	Sequence 239, App
863	25	1.8	1656	9	US-10-128-693A-239	Sequence 239, App
864	25	1.8	1656	9	US-10-131-813A-239	Sequence 239, App
865	25	1.8	1656	9	US-10-131-818A-239	Sequence 239, App
866	25	1.8	1656	9	US-10-131-823A-239	Sequence 239, App
867	25	1.8	1656	9	US-10-131-824A-239	Sequence 239, App
868	25	1.8	1656	9	US-10-131-830A-239	Sequence 239, App
869	25	1.8	1656	9	US-10-131-837A-239	Sequence 239, App
870	25	1.8	1656	9	US-10-137-872A-239	Sequence 239, App
871	25	1.8	1656	9	US-10-140-860-239	Sequence 239, App
872	25	1.8	1656	9	US-10-142-417-239	Sequence 239, App
873	25	1.8	1656	9	US-10-147-500-239	Sequence 239, App
874	25	1.8	1656	9	US-10-147-502-239	Sequence 239, App
875	25	1.8	1656	9	US-10-147-515-239	Sequence 239, App
876	25	1.8	1656	9	US-10-147-517-239	Sequence 239, App
877	25	1.8	1656	9	US-10-147-526-239	Sequence 239, App
878	25	1.8	1656	9	US-10-147-527-239	Sequence 239, App
879	25	1.8	1656	9	US-10-152-395-239	Sequence 239, App
880	25	1.8	1656	9	US-10-157-782-239	Sequence 239, App
881	25	1.8	1656	9	US-10-121-040-239	Sequence 239, App
882	25	1.8	1656	9	US-10-121-056-239	Sequence 239, App
883	25	1.8	1656	9	US-10-121-061-239	Sequence 239, App
884	25	1.8	1656	9	US-10-123-235-239	Sequence 239, App
885	25	1.8	1656	9	US-10-124-818-239	Sequence 239, App
886	25	1.8	1656	9	US-10-125-926A-239	Sequence 239, App
887	25	1.8	1656	9	US-10-125-930A-239	Sequence 239, App
888	25	1.8	1656	9	US-10-127-837A-239	Sequence 239, App
889	25	1.8	1656	9	US-10-127-838A-239	Sequence 239, App
890	25	1.8	1656	9	US-10-127-842A-239	Sequence 239, App
891	25	1.8	1656	9	US-10-127-843A-239	Sequence 239, App
892	25	1.8	1656	9	US-10-127-845A-239	Sequence 239, App
893	25	1.8	1656	9	US-10-127-846A-239	Sequence 239, App
894	25	1.8	1656	9	US-10-127-847A-239	Sequence 239, App
895	25	1.8	1656	9	US-10-127-848A-239	Sequence 239, App

969	25	1.8	1992	9	US-10-036-542-18	Sequence 18, Appl
C 970	25	1.8	2000	9	US-09-938-842A-3635	Sequence 3635, Ap
971	25	1.8	2010	10	US-09-809-545A-52	Sequence 52, Appl
972	25	1.8	2014	9	US-10-106-698-390	Sequence 390, Ap
973	25	1.8	2085	9	US-10-106-698-2105	Sequence 2105, Ap
974	25	1.8	2098	9	US-10-163-866-11	Sequence 11, Appl
C 975	25	1.8	2122	9	US-10-001-873-2	Sequence 1647, Ap
976	25	1.8	2124	9	US-10-106-698-1647	Sequence 13927, A
977	25	1.8	2209	9	US-10-198-846-13927	Sequence 110, App
978	25	1.8	2223	10	US-10-125-540-110	Sequence 160, App
979	25	1.8	2223	10	US-09-764-870-110	Sequence 152, App
980	25	1.8	2229	10	US-09-925-301-160	Sequence 7825, Ap
C 981	25	1.8	2230	10	US-09-822-849A-152	Sequence 1073, Ap
C 983	25	1.8	2295	9	US-09-764-891-7825	Sequence 311, App
984	25	1.8	2340	9	US-10-106-698-1073	Sequence 44, Appl
985	25	1.8	2342	10	US-09-925-297-311	Sequence 144, App
986	25	1.8	2359	9	US-10-091-438-44	Sequence 103, App
987	25	1.8	2370	10	US-09-764-853-144	Sequence 1, Appl
988	25	1.8	2391	9	US-09-822-846-103	Sequence 438, App
989	25	1.8	2402	9	US-10-194-125-1	Sequence 13, Appl
990	25	1.8	2438	9	US-10-102-806-13	Sequence 471, App
991	25	1.8	2557	10	US-09-925-300-471	Sequence 101, App
992	25	1.8	2574	9	US-09-978-295A-101	Sequence 101, App
993	25	1.8	2574	9	US-09-978-697-101	Sequence 101, App
994	25	1.8	2574	9	US-09-978-192A-101	Sequence 101, App
995	25	1.8	2574	9	US-09-999-832A-101	Sequence 101, App
996	25	1.8	2574	9	US-09-978-189-101	Sequence 101, App
997	25	1.8	2574	9	US-09-978-608A-101	Sequence 101, App
998	25	1.8	2574	9	US-09-978-191A-101	Sequence 101, App
999	25	1.8	2574	9	US-09-978-403A-101	Sequence 101, App
1000	25	1.8	2574	9	US-09-978-403A-101	Sequence 101, App

ALIGNMENTS

RESULT 1
US-09-989-919-15
Sequence 15, Application US/09989919
Patent No. US2002016434A1
GENERAL INFORMATION:
APPLICANT: Macina, Roberto
APPLICANT: Recipon, Hervé
APPLICANT: Pluta, Jason
APPLICANT: Ghosh, Malavika
APPLICANT: Sun, Yonming
APPLICANT: Liu, Chenghua
TITLE OF INVENTION: Compositions and Methods Relating to Colon Specific Genes and Pro
FILE REFERENCE: DEX-0289
CURRENT APPLICATION NUMBER: US/09/989,919
CURRENT FILING DATE: 2001-11-21
PRIOR APPLICATION NUMBER: 60/225,505
PRIOR FILING DATE: 2000-11-22
NUMBER OF SEQ ID NOS: 124
SOFTWARE: PatentIn version 3.1
SEQ ID NO 15
LENGTH: 1397
TYPE: DNA
ORGANISM: Homo sapien
US-09-989-919-15

Query Match 100.0%; Score 1397; DB 9; Length 1397;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1397; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 1 GATGCTGACCTGTATCCGAGGCGGAGTATCTGAGAACTCAAGGAGACAGACAGTAC 60
1 GGTGTGACCTGTATCCGAGGCGGAGTATCTGAGAACTCAAGGAGACAGACAGTAC 60
Qy 61 CGAGTACAGTGTATCCAGACAGACATCCCCAGAGAACTACGCTGTGCTCCTC 120
61 CGAGTACAGTGTATCCAGACAGACATCCCCAGAGAACTACGCTGTGCTCCTC 120

Qy 121 CTACACACAGGAGGAGCTGCTCTTTCAGTGTTCACCTGAGCTGAGTGTCTG 180
121 CTACACACAGGAGGAGCTGCTCTTTCAGTGTTCACCTGAGCTGAGTGTCTG 180
Db 121 CTACACACAGGAGGAGCTGCTCTTTCAGTGTTCACCTGAGCTGAGTGTCTG 180
Qy 181 TGAGAGCCATCCAGATGTGCGGCTTGTGTGTCACCAACCAACCTGAGAGTGA 240
181 TGAGAGCCATCCAGATGTGCGGCTTGTGTGTCACCAACCAACCTGAGAGTGA 240
Db 181 TGAGAGCCATCCAGATGTGCGGCTTGTGTGTCACCAACCAACCTGAGAGTGA 240
Qy 241 GCGAGTGGAGAGGCTTCCAGAGGAGATGCGAGACCTTCTGAGAGTGTATGATG 300
241 GCGAGTGGAGAGGCTTCCAGAGGAGATGCGAGACCTTCTGAGAGTGTATGATG 300
Db 241 GCGAGTGGAGAGGCTTCCAGAGGAGATGCGAGACCTTCTGAGAGTGTATGATG 300
Qy 301 TGATCCCCCATCGGAGTCAAGGAGGAGTGTGAGTGTATGAGAGAGTATGATG 360
301 TGATCCCCCATCGGAGTCAAGGAGGAGTGTGAGTGTATGAGAGAGTATGATG 360
Db 301 TGATCCCCCATCGGAGTCAAGGAGGAGTGTGAGTGTATGAGAGAGTATGATG 360
Qy 361 TCAGGAGTCAATTTAGGAGATGTGCTTCTCCAGAAAGAGAAATTCAGCCTTG 420
361 TCAGGAGTCAATTTAGGAGATGTGCTTCTCCAGAAAGAGAAATTCAGCCTTG 420
Db 361 TCAGGAGTCAATTTAGGAGATGTGCTTCTCCAGAAAGAGAAATTCAGCCTTG 420
Qy 421 TTACTCTTCACTGCTGCCCCCAGGTGCGAGCTGCTTTTTCAGAGCTGATGAGCC 480
421 TTACTCTTCACTGCTGCCCCCAGGTGCGAGCTGCTTTTTCAGAGCTGATGAGCC 480
Db 421 TTACTCTTCACTGCTGCCCCCAGGTGCGAGCTGCTTTTTCAGAGCTGATGAGCC 480
Qy 481 AGTGTGCTGATCCCAAGACACATATGTAGAGGCTGCTGATCTATCTGAG 540
481 AGTGTGCTGATCCCAAGACACATATGTAGAGGCTGCTGATCTATCTGAG 540
Db 481 AGTGTGCTGATCCCAAGACACATATGTAGAGGCTGCTGATCTATCTGAG 540
Qy 541 GCGTGGCTGACAGCTGATCTATCTGACAGCTGCTGCTGCTGAGAGAGTACTT 600
541 GCGTGGCTGACAGCTGATCTATCTGACAGCTGCTGCTGCTGAGAGAGTACTT 600
Db 541 GCGTGGCTGACAGCTGATCTATCTGACAGCTGCTGCTGCTGAGAGAGTACTT 600
Qy 601 GCACTGGAGCACTGATGTACCTGGAGACCTTCCAGACAACTGATTAACGACA 660
601 GCACTGGAGCACTGATGTACCTGGAGACCTTCCAGACAACTGATTAACGACA 660
Db 601 GCACTGGAGCACTGATGTACCTGGAGACCTTCCAGACAACTGATTAACGACA 660
Qy 661 GACAGATGACACAGACAAACGTCAATATGCAATTTAAATGATGATTAACG 720
661 GACAGATGACACAGACAAACGTCAATATGCAATTTAAATGATGATTAACG 720
Db 661 GACAGATGACACAGACAAACGTCAATATGCAATTTAAATGATGATTAACG 720
Qy 721 CCTAGTATGAGACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 780
721 CCTAGTATGAGACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 780
Db 721 CCTAGTATGAGACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 780
Qy 781 ACCGTGGGCTGTAGCAAGCTCAAGCTGCTGCTGCTGCTGCTGCTGCTGCT 840
781 ACCGTGGGCTGTAGCAAGCTCAAGCTGCTGCTGCTGCTGCTGCTGCTGCT 840
Db 781 ACCGTGGGCTGTAGCAAGCTCAAGCTGCTGCTGCTGCTGCTGCTGCTGCT 840
Qy 841 GGGACGGTTCGCTGGGAGCCCATCACTGTGTTCAATGCTGATGATGCTTAAC 900
841 GGGACGGTTCGCTGGGAGCCCATCACTGTGTTCAATGCTGATGATGCTTAAC 900
Db 841 GGGACGGTTCGCTGGGAGCCCATCACTGTGTTCAATGCTGATGATGCTTAAC 900
Qy 901 CCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 960
901 CCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 960
Db 901 CCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 960
Qy 961 TCGTGGAGTGTCTCTCAGCTTATGCTGAGACAGAGACTTGGCGGGAGTGTCCAGA 1020
961 TCGTGGAGTGTCTCTCAGCTTATGCTGAGACAGAGACTTGGCGGGAGTGTCCAGA 1020
Db 961 TCGTGGAGTGTCTCTCAGCTTATGCTGAGACAGAGACTTGGCGGGAGTGTCCAGA 1020
Qy 1021 TGTGGTATCTGTAATCTGGGAGGCTATCTGATCTCCGACAGGGGACATCCCG 1080
1021 TGTGGTATCTGTAATCTGGGAGGCTATCTGATCTCCGACAGGGGACATCCCG 1080
Db 1021 TGTGGTATCTGTAATCTGGGAGGCTATCTGATCTCCGACAGGGGACATCCCG 1080
Qy 1081 GCGAGCCAGGGTGTGAGGAGGAGGAGTGTGAGTGTGAGTGTGAGTGTGAGT 1140
1081 GCGAGCCAGGGTGTGAGGAGGAGGAGTGTGAGTGTGAGTGTGAGTGTGAGT 1140
Db 1081 GCGAGCCAGGGTGTGAGGAGGAGGAGTGTGAGTGTGAGTGTGAGTGTGAGT 1140
Qy 1141 GGAGCAGTGTGTTTGTGAGCAGAGACTTGGGGGCGGGGCTTCTGCTCT 1200
1141 GGAGCAGTGTGTTTGTGAGCAGAGACTTGGGGGCGGGGCTTCTGCTCT 1200
Db 1141 GGAGCAGTGTGTTTGTGAGCAGAGACTTGGGGGCGGGGCTTCTGCTCT 1200

QY 1201 CATTGCTTCATGAAAGCTCAAGACCAAAACAGGCTTTCCCTCTCTGAGT 1260
DB 1201 CATTGCTTCATGAAAGCTCAAGACCAAAACAGGCTTTCCCTCTCTGAGT 1260
QY 1261 TTGAATATCAGATCTTTGTACTCTTGTGTAAATGTTATTTTGTAAAAAT 1320
DB 1261 TTGAATATCAGATCTTTGTACTCTTGTGTAAATGTTATTTTGTAAAAAT 1320
QY 1321 AAAATAAATTAATTAATAAATGATGTTTACAGCAAACTCTTCCCTAAAAA 1380
DB 1321 AAAATAAATTAATTAATAAATGATGTTTACAGCAAACTCTTCCCTAAAAA 1380
QY 1381 AAAAAAAAAAGCGGCTC 1397
DB 1381 AAAAAAAAAAGCGGCTC 1397

RESULT 2
US-09-989-919-14
; Sequence 14, Application US/09989919
; Patent No. US20020164344A1
; GENERAL INFORMATION:
; APPLICANT: Macina, Roberto
; APPLICANT: Racipon, Hervé
; APPLICANT: Pluta, Jason
; APPLICANT: Ghosh, Malavika
; APPLICANT: Sun, Yongming
; APPLICANT: Liu, Chenghua
; TITLE OF INVENTION: Compositions and Methods Relating to Colon Specific Genes and Pro
; FILE REFERENCE: DEX-0289
; CURRENT APPLICATION NUMBER: US/09/989,919
; CURRENT FILING DATE: 2001-11-21
; PRIOR APPLICATION NUMBER: 60/252,505
; PRIOR FILING DATE: 2000-11-22
; NUMBER OF SEQ ID NOS: 124
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 14
; LENGTH: 470
; TYPE: DNA
; ORGANISM: Homo sapien
US-09-989-919-14

Query Match 31.6%; Score 442; DB 9; Length 470;
Best Local Similarity 100.0%; Pred. No. 6,8e-211;
Matches 442; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 572 GCTGGGCTTCCCTGTGAGGAGTGAATGCTGCTGAGCAGACATGTAATCTGGGAAC 631
DB 7 GCTGGGCTTCCCTGTGAGGAGTGAATGCTGCTGAGCAGACATGTAATCTGGGAAC 66
QY 632 CCTGAGACAAAGCTTAACATCCAGACAGACAGATGTGACAGACAAACGTGCAATA 691
DB 67 CCTGAGACAAAGCTTAACATCCAGACAGACAGATGTGACAGACAAACGTGCAATA 126
QY 692 TGGCAATGTTAAATGTGATTTACAGCTTACAGCTATGAGATGCTGCTCTCTAGTCA 751
DB 127 TGGCAATGTTAAATGTGATTTACAGCTTACAGCTATGAGATGCTGCTCTAGTCA 186
QY 752 GGAATCATGGGGATGATGCTGCTCTCAACCTGTGGGCTGTAAAGCAAGCTCAGCTAG 811
DB 187 GGAATCATGGGGATGATGCTGCTCTCAACCTGTGGGCTGTAAAGCAAGCTCAGCTAG 246
QY 812 TCTCCCACTGGGGGCTGTGCTCCCTGCTGAGGAGGCTTCCGTGGGAGCCCATCACTGT 871
DB 247 TCTCCCACTGGGGGCTGTGCTCCCTGCTGAGGAGGCTTCCGTGGGAGCCCATCACTGT 306
QY 872 GTTCAATAGTGTGAGAAATGTAAGCCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 931
DB 307 GTTCAATAGTGTGAGAAATGTAAGCCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 366
QY 932 GGGGTGGGGGCTGCTGGGGGCAATCCAGCTGAGATGTTCTTCACTAGCTTAGTCTGGA 991
DB 367 GGGGTGGGGGCTGCTGGGGGCAATCCAGCTGAGATGTTCTTCACTAGCTTAGTCTGGA 426

QY 992 CAGAGACTTGGCGGGGATGC 1013
DB 427 CAGAGACTTGGCGGGGATGC 448

RESULT 3
US-09-880-107-1138/c
; Sequence 1138, Application US/09880107
; Patent No. US20020142981A1
; GENERAL INFORMATION:
; APPLICANT: Horne, Darci T.
; APPLICANT: Vockley, Joseph G.
; APPLICANT: Scherf, Uwe
; APPLICANT: Gene Logic, Inc.
; TITLE OF INVENTION: Gene Expression Profiles in Liver Cancer
; FILE REFERENCE: 44921-5028-WO
; CURRENT APPLICATION NUMBER: US/09/880,107
; CURRENT FILING DATE: 2001-06-14
; PRIOR APPLICATION NUMBER: US 60/211,379
; PRIOR FILING DATE: 2000-06-14
; PRIOR APPLICATION NUMBER: US 60/237,054
; PRIOR FILING DATE: 2000-10-02
; NUMBER OF SEQ ID NOS: 3950
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1138
; LENGTH: 427
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: Genbank Accession No. US20020142981A1 AA451877
US-09-880-107-1138

Query Match 23.0%; Score 322; DB 10; Length 427;
Best Local Similarity 99.5%; Pred. No. 7.1e-151;
Matches 422; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 936 GTGGGGCTGCTGGGGGCAATCCATCCGATGTTCTCTCAGCTTAGTCTGACAG 995
DB 427 GTGGGGCTGCTGGGGGCAATCCATCCGATGTTCTCTCAGCTTAGTCTGACAG 368
QY 996 AGACTTGGCGGGGATGCTCCAGATGTGGTGAATTTCTTACCTGGGGAGCTATCTG 1055
DB 367 AGACTTGGCGGGGATGCTCCAGATGTGGTGAATTTCTTACCTGGGGAGCTATCTG 308
QY 1056 ACCCTCCGACAGGGACAATCCAGGCGCCAGGCGCTCAGGGGAGAGGTGCACACT 1115
DB 307 ACCCTCCGACAGGGACAATCCAGGCGCCAGGCGCTCAGGGGAGAGGTGCACACT 248
QY 1116 CAGCATGAGCCAAAGCTGGGGGTCAAGGAGCAGGTGTGTTGAGCCAGGACCTGGGGCGG 1175
DB 247 CAGCATGAGCCAAAGCTGGGGGTCAAGGAGCAGGTGTGTTGAGCCAGGACCTGGGGCGG 188
QY 1176 GGGTGGGCGGGGCTTCTGCTCTCAATTTGCTTTCAATGAAGCTCAAGACAGCAAA 1235
DB 187 GGGTGGGCGGGGCTTCTGCTCTCAATTTGCTTTCAATGAAGCTCAAGACAGCAAA 128
QY 1236 ACCAGGCTTCCCTCTCTGCAATTTGAATTCAGAAATCTTTTGTACTTCTGTTGT 1295
DB 127 ACCAGGCTTCCCTCTCTGCAATTTGAATTCAGAAATCTTTTGTACTTCTGTTGT 68
QY 1296 TAAATGTTATTTTGTAAAAATTAATAAATAGTTAATAAATAGTTTCAAG 1355
DB 67 TAAATGTTATTTTGTAAAAATTAATAAATAGTTAATAAATAGTTTCAAG 8
QY 1356 CAAA 1359
DB 7 CAAA 4

RESULT 4
US-09-918-995-32213
; Sequence 32213, Application US/09918995

Publication No. US20030073623A1
GENERAL INFORMATION:
APPLICANT: Hyseq, Inc.
TITLE OF INVENTION: NOVEL NUCLEIC ACID SEQUENCES OBTAINED
FROM VARIOUS CDNA LIBRARIES
FILE REFERENCE: 20411-756
CURRENT APPLICATION NUMBER: US/09/918,995
CURRENT FILING DATE: 2001-07-30
PRIOR APPLICATION NUMBER: US/09/235,076
PRIOR FILING DATE: 1999-01-20
NUMBER OF SEQ ID NOS: 38054
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 32213
LENGTH: 493
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: misc_feature
LOCATION: (1)...(493)
OTHER INFORMATION: n = A,T,C or G
US-09-918-995-32213

Query Match 6.9%; Score 97; DB 9; Length 493;
Best Local Similarity 100.0%; Pred. No. 2,4e-38;
Matches 97; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GGTCGTGACCTGTACCGAGCGGAGCATCTGCAGAACTCCACGGCAGCAGCACTAC 60
DB 397 GGTCGTGACCTGTACCGAGCGGAGCATCTGCAGAACTCCACGGCAGCAGCACTAC 456
DB 457 CGAGTACCAGTGTATCCCGACAGACCATCCCCCAG 97
CGAGTACCAGTGTATCCCGACAGACCATCCCCCAG 493

RESULT 5
US-09-770-791-70/c
Sequence 70, Application US/0970791
Patent No. US20020062014A1
GENERAL INFORMATION:

APPLICANT: Gorlach, Jorn
APPLICANT: An, Yong-Qiang
APPLICANT: Hamilton, Carol M.
APPLICANT: Price, Jennifer L.
APPLICANT: Raines, Tracy M.
APPLICANT: Yu, Yang
APPLICANT: Rameaka, Joshua G.
APPLICANT: Page, Amy
APPLICANT: Mathew, Abraham V.
APPLICANT: Ledford, Brooke L.
APPLICANT: Woessner, Jeffrey P.
APPLICANT: Haas, William David
APPLICANT: Garcia, Carlos A.
APPLICANT: Kriker, Maja
APPLICANT: Slader, Ted
APPLICANT: Davis, Keith R.
APPLICANT: Allen, Keith
APPLICANT: Hoffman, Neil
APPLICANT: Hurban, Patrick
TITLE OF INVENTION: Expressed Sequences of Arabidopsis
TITLE OF INVENTION: Chailana
FILE REFERENCE: 2029 (PAPA-018PRV)
CURRENT APPLICATION NUMBER: US/09/770,791
CURRENT FILING DATE: 2001-01-26
PRIOR APPLICATION NUMBER: 60/178,480
PRIOR FILING DATE: 2000-01-27
NUMBER OF SEQ ID NOS: 999
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 70
LENGTH: 386
TYPE: DNA
ORGANISM: Arabidopsis thaliana
US-09-770-791-70

Query Match 2.1%; Score 30; DB 10; Length 386;
Best Local Similarity 100.0%; Pred. No. 8e-05;
Matches 30; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1361 TCTTCCCTTAAAAAAAAAAAAAAAAAAAA 1390
DB 31 TCTTCCCTTAAAAAAAAAAAAAAAAAAAA 2

RESULT 6
US-09-867-701-3641/c
Sequence 3641, Application US/09867701
Patent No. US20020132237A1
GENERAL INFORMATION:

APPLICANT: Aglate, Paul A.
APPLICANT: Jones, Robert
APPLICANT: Harlocker, Susan L.
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
FILE REFERENCE: 210121.497
CURRENT APPLICATION NUMBER: US/09/867,701
CURRENT FILING DATE: 2001-05-29
NUMBER OF SEQ ID NOS: 10912
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 3641
LENGTH: 421
TYPE: DNA
ORGANISM: Homo sapien
US-09-867-701-3641

Query Match 2.1%; Score 30; DB 10; Length 421;
Best Local Similarity 100.0%; Pred. No. 8e-05;
Matches 30; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1361 TCTTCCCTTAAAAAAAAAAAAAAAAAAAA 1390
DB 30 TCTTCCCTTAAAAAAAAAAAAAAAAAAAA 1

RESULT 7
US-09-770-445-884/c
Sequence 884, Application US/0970445
Patent No. US20020023281A1
GENERAL INFORMATION:

APPLICANT: Gorlach, Jorn
APPLICANT: An, Yong-Qiang
APPLICANT: Hamilton, Carol M.
APPLICANT: Price, Jennifer L.
APPLICANT: Raines, Tracy M.
APPLICANT: Yu, Yang
APPLICANT: Rameaka, Joshua G.
APPLICANT: Page, Amy
APPLICANT: Mathew, Abraham V.
APPLICANT: Ledford, Brooke L.
APPLICANT: Woessner, Jeffrey P.
APPLICANT: Haas, William David
APPLICANT: Garcia, Carlos A.
APPLICANT: Kriker, Maja
APPLICANT: Slader, Ted
APPLICANT: Davis, Keith R.
APPLICANT: Allen, Keith
APPLICANT: Hoffman, Neil
APPLICANT: Hurban, Patrick
TITLE OF INVENTION: Expressed Sequences of Arabidopsis
TITLE OF INVENTION: Chailana
FILE REFERENCE: 2023US (PAPA-012PRV)
CURRENT APPLICATION NUMBER: US/09/770,445
CURRENT FILING DATE: 2001-01-26
PRIOR APPLICATION NUMBER: US 60/178,472
PRIOR FILING DATE: 2000-01-27
NUMBER OF SEQ ID NOS: 999
SOFTWARE: FastSeq for Windows Version 4.0

SEQ ID NO 884.
LENGTH: 777
TYPE: DNA
ORGANISM: Arabidopsis thaliana
US-09-770-445-884

Query Match
Best Local Similarity 100.0%; Pred. No. 7.9e-05;
Matches 30; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1361 TCTTCCTTAAAAAAAAAAAAAAAAAAAA 1390
DB 45 TCTTCCTTAAAAAAAAAAAAAAAAAAAA 16

RESULT 8
US-10-174-590-569
Sequence 569, Application US/10174590
Publication No. US20030008352A1

GENERAL INFORMATION:
APPLICANT: Baker, Kevin P.
APPLICANT: Chen, Jian
APPLICANT: Desnoyers, Luc
APPLICANT: Goddard, Audrey
APPLICANT: Godowski, Paul J.
APPLICANT: Gurney, Austin L.
APPLICANT: Pan, James
APPLICANT: Smith, Victoria
APPLICANT: Watanabe, Colin K.
APPLICANT: Wood, William I.
APPLICANT: Zhang, Zemin
TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
FILE REFERENCE: P3430R1C42
CURRENT APPLICATION NUMBER: US/10/174,590
CURRENT FILING DATE: 2002-06-18
Prior application removed - See File Wrapper or Palm
NUMBER OF SEQ ID NOS: 612
SEQ ID NO 569
LENGTH: 2457
TYPE: DNA
ORGANISM: Homo Sapien
US-10-174-590-569

Query Match
Best Local Similarity 100.0%; Pred. No. 7.7e-05;
Matches 30; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1361 TCTTCCTTAAAAAAAAAAAAAAAAAAAA 1390
DB 2423 TCTTCCTTAAAAAAAAAAAAAAAAAAAA 2452

RESULT 9
US-10-176-758-569
Sequence 569, Application US/10176758
Publication No. US20030008353A1

GENERAL INFORMATION:
APPLICANT: Baker, Kevin P.
APPLICANT: Chen, Jian
APPLICANT: Desnoyers, Luc
APPLICANT: Goddard, Audrey
APPLICANT: Godowski, Paul J.
APPLICANT: Gurney, Austin L.
APPLICANT: Pan, James
APPLICANT: Smith, Victoria
APPLICANT: Watanabe, Colin K.
APPLICANT: Wood, William I.
APPLICANT: Zhang, Zemin
TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
FILE REFERENCE: P3430R1C104
CURRENT APPLICATION NUMBER: US/10/176,758

CURRENT FILING DATE: 2002-06-21
Prior Application removed - See File Wrapper or Palm
NUMBER OF SEQ ID NOS: 612
SEQ ID NO 569
LENGTH: 2457
TYPE: DNA
ORGANISM: Homo Sapien
US-10-176-758-569

Query Match
Best Local Similarity 100.0%; Pred. No. 7.7e-05;
Matches 30; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1361 TCTTCCTTAAAAAAAAAAAAAAAAAAAA 1390
DB 2423 TCTTCCTTAAAAAAAAAAAAAAAAAAAA 2452

RESULT 10
US-10-175-737-569
Sequence 569, Application US/10175737
Publication No. US20030013153A1

GENERAL INFORMATION:
APPLICANT: Baker, Kevin P.
APPLICANT: Chen, Jian
APPLICANT: Desnoyers, Luc
APPLICANT: Goddard, Audrey
APPLICANT: Godowski, Paul J.
APPLICANT: Gurney, Austin L.
APPLICANT: Pan, James
APPLICANT: Smith, Victoria
APPLICANT: Watanabe, Colin K.
APPLICANT: Wood, William I.
APPLICANT: Zhang, Zemin
TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
FILE REFERENCE: P3430R1C50
CURRENT APPLICATION NUMBER: US/10/175,737
CURRENT FILING DATE: 2002-06-19
Prior Application removed - See File Wrapper or Palm
NUMBER OF SEQ ID NOS: 612
SEQ ID NO 569
LENGTH: 2457
TYPE: DNA
ORGANISM: Homo Sapien
US-10-175-737-569

Query Match
Best Local Similarity 100.0%; Pred. No. 7.7e-05;
Matches 30; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1361 TCTTCCTTAAAAAAAAAAAAAAAAAAAA 1390
DB 2423 TCTTCCTTAAAAAAAAAAAAAAAAAAAA 2452

RESULT 11
US-10-173-706-569
Sequence 569, Application US/10173706
Publication No. US2003002293A1

GENERAL INFORMATION:
APPLICANT: Baker, Kevin P.
APPLICANT: Chen, Jian
APPLICANT: Desnoyers, Luc
APPLICANT: Goddard, Audrey
APPLICANT: Godowski, Paul J.
APPLICANT: Gurney, Austin L.
APPLICANT: Pan, James
APPLICANT: Smith, Victoria
APPLICANT: Watanabe, Colin K.
APPLICANT: Wood, William I.
APPLICANT: Zhang, Zemin
TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC

TITLE OF INVENTION: ACIDS ENCODING THE SAME
FILE REFERENCE: P3430R1C7
CURRENT APPLICATION NUMBER: US/10/173,706
CURRENT FILING DATE: 2002-06-17
Prior Application removed - See File Wrapper or Palm
NUMBER OF SEQ ID NOS: 612
SEQ ID NO 569
LENGTH: 2457
TYPE: DNA
ORGANISM: Homo Sapien
US-10-173-706-569

Query Match 2.1%; Score 30; DB 9; Length 2457;
Best Local Similarity 100.0%; Pred. No. 7.7e-05;
Matches 30; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1361 TCTTCCCTAATAAAAAAAAAAAAAAAAAA 1390
DB 2423 TCTTCCCTAATAAAAAAAAAAAAAAAAAA 2452

RESULT 12

US-10-175-738-569
Sequence 569, Application US/10175738
Publication No. US20030022294A1

GENERAL INFORMATION:
APPLICANT: Baker, Kevin P.
APPLICANT: Chen, Jian
APPLICANT: Desnoyers, Luc
APPLICANT: Goddard, Audrey
APPLICANT: Godowski, Paul J.
APPLICANT: Gurney, Austin L.
APPLICANT: Pan, James
APPLICANT: Smith, Victoria
APPLICANT: Watanabe, Colin K.
APPLICANT: Wood, William I.
APPLICANT: Zhang, Zemin
TITLE OF INVENTION: ACIDS ENCODING THE SAME
FILE REFERENCE: P3430R1C45
CURRENT APPLICATION NUMBER: US/10/175,738
CURRENT FILING DATE: 2002-06-19
Prior Application removed - See File Wrapper or Palm
NUMBER OF SEQ ID NOS: 612
SEQ ID NO 569
LENGTH: 2457
TYPE: DNA
ORGANISM: Homo Sapien
US-10-175-738-569

Query Match 2.1%; Score 30; DB 9; Length 2457;
Best Local Similarity 100.0%; Pred. No. 7.7e-05;
Matches 30; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1361 TCTTCCCTAATAAAAAAAAAAAAAAAAAA 1390
DB 2423 TCTTCCCTAATAAAAAAAAAAAAAAAAAA 2452

RESULT 13

US-10-175-752-569
Sequence 569, Application US/10175752
Publication No. US20030022295A1

GENERAL INFORMATION:
APPLICANT: Baker, Kevin P.
APPLICANT: Chen, Jian
APPLICANT: Desnoyers, Luc
APPLICANT: Goddard, Audrey
APPLICANT: Godowski, Paul J.
APPLICANT: Gurney, Austin L.
APPLICANT: Pan, James
APPLICANT: Smith, Victoria
APPLICANT: Watanabe, Colin K.

APPLICANT: Wood, William I.
APPLICANT: Zhang, Zemin
TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
FILE REFERENCE: P3430R1C60
CURRENT APPLICATION NUMBER: US/10/175,752
CURRENT FILING DATE: 2002-06-19
Prior Application removed - See File Wrapper or Palm
NUMBER OF SEQ ID NOS: 612
SEQ ID NO 569
LENGTH: 2457
TYPE: DNA
ORGANISM: Homo Sapien
US-10-175-752-569

Query Match 2.1%; Score 30; DB 9; Length 2457;
Best Local Similarity 100.0%; Pred. No. 7.7e-05;
Matches 30; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1361 TCTTCCCTAATAAAAAAAAAAAAAAAAAA 1390
DB 2423 TCTTCCCTAATAAAAAAAAAAAAAAAAAA 2452

RESULT 14

US-10-176-482-569
Sequence 569, Application US/10176482
Publication No. US20030022296A1

GENERAL INFORMATION:
APPLICANT: Baker, Kevin P.
APPLICANT: Chen, Jian
APPLICANT: Desnoyers, Luc
APPLICANT: Goddard, Audrey
APPLICANT: Godowski, Paul J.
APPLICANT: Gurney, Austin L.
APPLICANT: Pan, James
APPLICANT: Smith, Victoria
APPLICANT: Watanabe, Colin K.
APPLICANT: Wood, William I.
APPLICANT: Zhang, Zemin
TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
FILE REFERENCE: P3430R1C70
CURRENT APPLICATION NUMBER: US/10/176,482
CURRENT FILING DATE: 2002-06-20
Prior Application removed - See File Wrapper or Palm
NUMBER OF SEQ ID NOS: 612
SEQ ID NO 569
LENGTH: 2457
TYPE: DNA
ORGANISM: Homo Sapien
US-10-176-482-569

Query Match 2.1%; Score 30; DB 9; Length 2457;
Best Local Similarity 100.0%; Pred. No. 7.7e-05;
Matches 30; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1361 TCTTCCCTAATAAAAAAAAAAAAAAAAAA 1390
DB 2423 TCTTCCCTAATAAAAAAAAAAAAAAAAAA 2452

RESULT 15

US-10-176-757-569
Sequence 569, Application US/10176757
Publication No. US20030022297A1

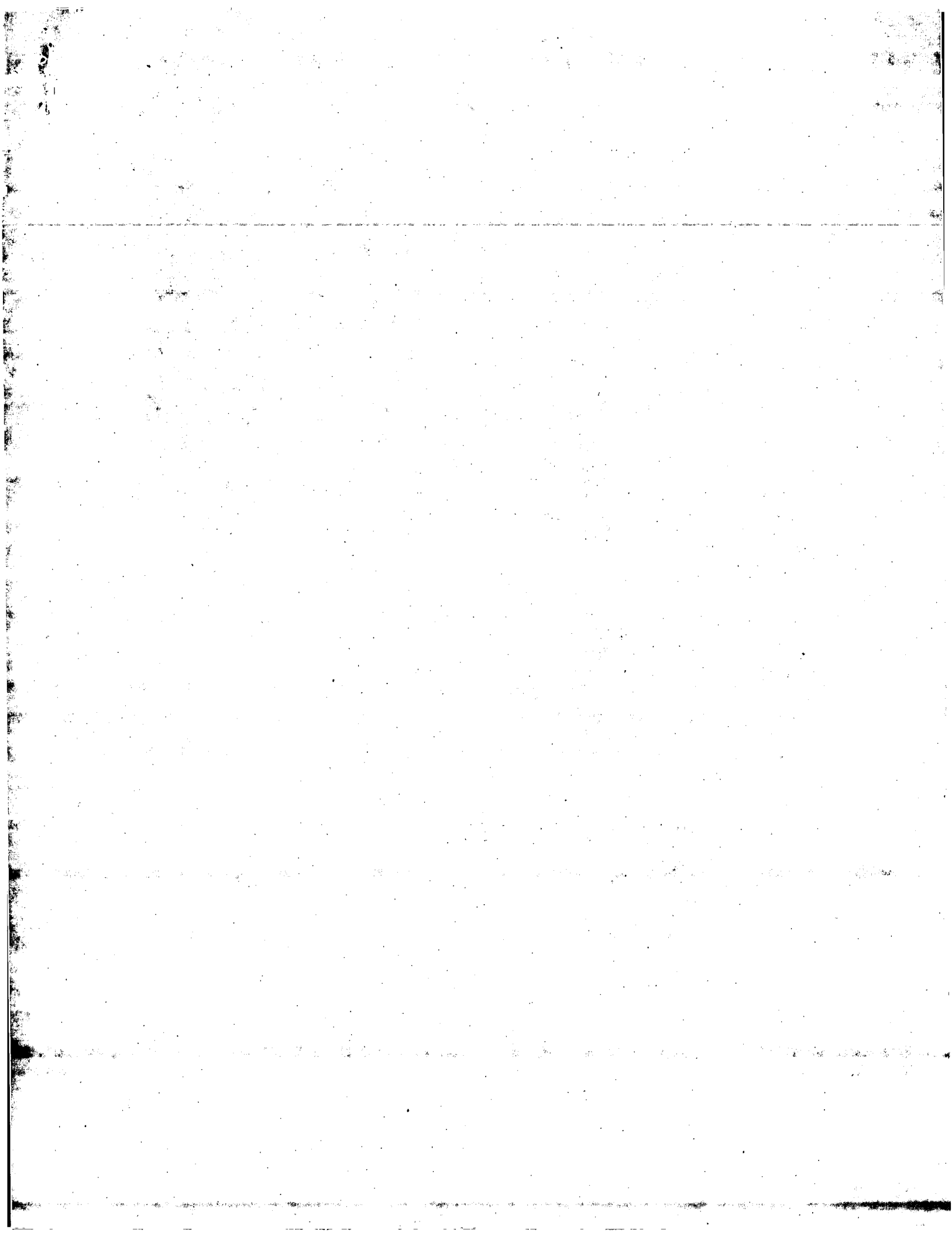
GENERAL INFORMATION:
APPLICANT: Baker, Kevin P.
APPLICANT: Chen, Jian
APPLICANT: Desnoyers, Luc
APPLICANT: Goddard, Audrey
APPLICANT: Godowski, Paul J.
APPLICANT: Gurney, Austin L.

; APPLICANT: Pan,James
 ; APPLICANT: Smith,Victoria
 ; APPLICANT: Watanabe,Colin K.
 ; APPLICANT: Wood,William I.
 ; APPLICANT: Zhang,Zemin
 ; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
 ; FILE OF INVENTION: ACIDS ENCODING THE SAME
 ; FILE REFERENCE: P3430R1C86
 ; CURRENT APPLICATION NUMBER: US/10/176,757
 ; CURRENT FILING DATE: 2002-06-20
 ; Prior Application removed - See File Wrapper or Palm
 ; NUMBER OF SEQ ID NOS: 612
 ; SEQ ID NO 569
 ; LENGTH: 2457
 ; TYPE: DNA
 ; ORGANISM: Homo Sapien
 US-10-176-757-569

Query Match 2.1%; Score 30; DB 9; Length 2457;
 Best Local Similarity 100.0%; Pred. No. 7.7e-05;
 Matches 30; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1361 TCTTCCTAAAAAAAAAAAAAAAAAAAA 1390
 DB 2423 TCTTCCTAAAAAAAAAAAAAAAAAAAA 2452

Search completed: June 29, 2003, 06:40:51
 Job time: 185 secs



GenCore version 5.1.6
Copyright (c) 1993 - 2003 Comugen Ltd.

OM nucleic - nucleic search, using sw-model

Run on: June 29, 2003, 05:21:29 ; Search time 1461 Seconds
(without alignments)
15486.034 Million cell updates/sec

Title: US-09-989-919-15

Perfect score: 1397

Sequence: 1 ggtgctgcacctgtaccgga.....aaaaaaaaaaagcggtc 1397

Scoring table: OLIGO NUC

Gapop 60.0 , Gapext 60.0

Searched: 16154066 seqs, 8097743376 residues

Word size : 0

Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 1000 summaries

Database :
EST:
1: em_escba:*
2: em_escbum:*
3: em_escin:*
4: em_escmu:*
5: em_escrov:*
6: em_escpl:*
7: em_escro:*
8: em_esci:*
9: gb_esc1:*
10: gb_esc2:*
11: gb_hcc:*
12: gb_esc3:*
13: gb_esc4:*
14: gb_esc5:*
15: em_escfun:*
16: em_escrom:*
17: gb_gss:*
18: em_gss_hum:*
19: em_gss_inv:*
20: em_gss_vln:*
21: em_gss_vrt:*
22: em_gss_fun:*
23: em_gss_mam:*
24: em_gss_mus:*
25: em_gss_oher:*
26: em_gss_pro:*
27: em_gss_rod:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	587	42.0	1039	14	BM9227766 AGENCOURT
2	566	40.5	776	12	BG827019 602749053
3	520	37.2	520	14	BO013425 UI-1-BCIP
4	517	37.0	529	9	A1669816 UI-1-BCIP
5	514	36.8	770	10	BE250262 600943273
6	513	36.7	773	14	BM671384 UI-E-CX1

7	493	35.3	610	13	BM668538 UI-E-CX1
8	492	35.2	543	13	BM142660 1926C02.x
9	486	34.8	504	10	BE677292 7d8b01.x
10	486	34.8	543	14	BM726293 UI-E-EJO
11	478	34.2	492	9	A1453409 UI-E-EJO
12	473	33.9	485	14	BQ012938 UI-E-EJO
13	471	33.7	522	14	BQ888182 AGENCOURT
14	467	33.4	823	10	AM592430 h44e03.x
15	465	33.3	465	14	BM693225 UI-E-CX1
16	463	33.1	565	10	AM516922 xp88g09.x
17	448	32.1	548	14	BM692817 UI-E-CX1
18	445	31.9	493	10	AM263720 UI-E-CX1
19	442	31.6	493	10	BE205791 UI-E-CX1
20	431	30.9	494	9	AA552700 UI-E-CX1
21	430	30.8	481	10	AM027861 UI-E-CX1
22	429	30.7	419	9	A1357244 UI-E-CX1
23	419	30.0	419	10	AM167704 UI-E-CX1
24	418	29.9	504	9	A1808803 UI-E-CX1
25	412	29.5	419	14	BQ270709 UI-E-CX1
26	412	29.5	434	9	A1676190 UI-E-CX1
27	409	29.3	455	14	BQ270467 UI-E-CX1
28	407	29.1	504	10	AM411057 UI-E-CX1
29	403	28.8	454	10	AM058133 UI-E-CX1
30	403	28.8	999	14	BQ068821 AGENCOURT
31	401	28.7	452	9	A1694266 UI-E-CX1
32	397	28.4	458	10	AM518556 UI-E-CX1
33	396	28.3	498	10	AM512747 UI-E-CX1
34	390	27.9	645	13	BM665930 UI-E-CX1
35	386	27.6	444	14	BM689301 UI-E-CX1
36	384	27.5	435	10	AM026902 UI-E-CX1
37	384	27.5	435	10	BE502518 UI-E-CX1
38	382	27.3	433	10	AM268497 UI-E-CX1
39	374	26.8	429	14	BM854794 UI-E-CX1
40	370	26.5	421	9	AA936915 UI-E-CX1
41	369	26.4	420	10	AM469401 UI-E-CX1
42	366	26.2	366	9	A1336211 UI-E-CX1
43	364	26.1	665	12	BF528878 UI-E-CX1
44	361	25.8	361	9	AA620861 UI-E-CX1
45	356	25.5	421	10	AM304245 UI-E-CX1
46	353	25.3	394	10	AM089064 UI-E-CX1
47	351	25.1	394	9	AA722927 UI-E-CX1
48	351	25.1	397	14	BM714252 UI-E-CX1
49	351	25.1	400	14	BM676630 UI-E-CX1
50	351	25.1	434	9	A1375823 UI-E-CX1
51	351	25.1	468	9	A1342651 UI-E-CX1
52	350	25.1	678	10	AM409897 UI-E-CX1
53	345	24.7	345	12	BF17971 UI-E-CX1
54	342	24.5	342	10	AM517777 UI-E-CX1
55	341	24.4	341	9	AA927392 UI-E-CX1
56	340	24.3	352	9	A1193827 UI-E-CX1
57	325	23.3	512	10	AM960467 UI-E-CX1
58	322	23.0	427	9	AA451877 UI-E-CX1
59	322	23.0	521	14	BM680899 UI-E-CX1
60	321	23.0	414	10	AM455766 UI-E-CX1
61	319	22.8	888	14	BO652861 UI-E-CX1
62	318	22.8	917	14	BO654344 UI-E-CX1
63	317	22.7	419	9	AA631015 UI-E-CX1
64	317	22.7	489	10	AM001523 UI-E-CX1
65	306	21.9	490	9	A1581877 UI-E-CX1
66	281	20.1	433	10	AM592074 UI-E-CX1
67	280	20.0	456	9	A1539296 UI-E-CX1
68	279	20.0	518	9	A1131395 UI-E-CX1
69	273	19.5	545	14	W19150 UI-E-CX1
70	272	19.5	424	9	AA430096 UI-E-CX1
71	269	19.3	1038	14	EM927767 UI-E-CX1
72	267	19.1	277	9	AA877453 UI-E-CX1
73	263	18.8	315	9	AA913369 UI-E-CX1
74	263	18.8	416	10	AM517168 UI-E-CX1
75	259	18.5	362	10	AA643852 UI-E-CX1
76	253	18.1	462	9	A1355782 UI-E-CX1
77	251	18.0	251	9	A1219633 UI-E-CX1
78	249	17.8	249	9	BM738028 UI-E-CX1
79	249	17.8	568	14	BM738028 UI-E-CX1

226	29	2.1	427	9	A1659585	A1659585 t87c06.x	C 299	28	2.0	334	9	A1659138	A1659138 t87c06.x
227	29	2.1	442	10	AM667554	AM667554 GA_Ea000	C 300	28	2.0	341	9	A1379081	A1379081 t87c06.x
228	29	2.1	450	9	A1537465	A1537465 t87c06.x	C 301	28	2.0	343	9	A1360535	A1360535 t87c06.x
229	29	2.1	480	14	B0523753	B0523753 NISC_n124	C 302	28	2.0	344	10	AM058351	AM058351 t87c06.x
230	29	2.1	488	9	AA048125	AA048125 m127d03.r	C 303	28	2.0	345	10	AM058351	AM058351 t87c06.x
231	29	2.1	486	9	A1332759	A1332759 qp92e08.x	C 304	28	2.0	347	9	A1552641	A1552641 t87c06.x
232	29	2.1	533	9	A1326071	A1326071 m72a11.x	C 305	28	2.0	347	9	A1552641	A1552641 t87c06.x
233	29	2.1	558	14	B0060374	B0060374 Ts1445.Th	C 306	28	2.0	356	12	BF407940	BF407940 UI-R-BJ2
234	29	2.1	570	9	AL589617	AL589617 DKE5451M	C 307	28	2.0	356	13	BM254936	BM254936 516522.MA
235	29	2.1	640	9	A1676094	A1676094 W05C09.x	C 308	28	2.0	358	13	BM1845732	BM1845732 t87c06.x
236	29	2.1	761	14	B0792946	B0792946 EST_7851	C 309	28	2.0	360	13	BG921039	BG921039 602822963
237	28	2.0	100	14	B0268655	B0268655 1156f09.x	C 310	28	2.0	366	14	R43409	R43409 y908f01.s1
238	28	2.0	103	12	BF910807	BF910807 MRO-UT004	C 311	28	2.0	371	9	AA769422	AA769422 n238512.s
239	28	2.0	104	12	BF936908	BF936908 EST45963	C 312	28	2.0	372	10	AM288995	AM288995 707002H02
240	28	2.0	112	9	A1247732	A1247732 QH56502.x	C 313	28	2.0	376	9	A1538908	A1538908 t87c06.x
241	28	2.0	120	9	A1560545	A1560545 t48907.x	C 314	28	2.0	381	10	AM487692	AM487692 t87c06.x
242	28	2.0	139	12	BG173790	BG173790 60233384	C 315	28	2.0	382	10	AM167215	AM167215 t87c06.x
243	28	2.0	146	9	A1793210	A1793210 q231g01.x	C 316	28	2.0	382	13	BM381348	BM381348 t87c06.x
244	28	2.0	146	12	BG793028	BG793028 TWSW_H7P3	C 317	28	2.0	398	13	B1118835	B1118835 t87c06.x
245	28	2.0	168	12	BG779691	BG779691 602668510	C 318	28	2.0	400	9	AA831172	AA831172 oc73p04.s
246	28	2.0	169	9	AU052628	AU052628 AU052628	C 319	28	2.0	408	10	AM037241	AM037241 614021H11
247	28	2.0	169	12	BE959472	BE959472 601654411	C 320	28	2.0	409	13	BM169498	BM169498 t87c06.x
248	28	2.0	173	10	AM024712	AM024712 w068905.x	C 321	28	2.0	410	9	A1473375	A1473375 t87c06.x
249	28	2.0	174	12	BG359964	BG359964 d4b8902.	C 322	28	2.0	412	9	A1072964	A1072964 UI-R-Y0-1
250	28	2.0	182	9	AU038663	AU038663 w019d05.x	C 323	28	2.0	414	9	A1963656	A1963656 w764c09.x
251	28	2.0	186	13	BG977864	BG977864 MR2-C1012	C 324	28	2.0	415	10	BE477186	BE477186 160772.BA
252	28	2.0	190	13	BM026529	BM026529 f106d06.x	C 325	28	2.0	416	13	BM378314	BM378314 t87c06.x
253	28	2.0	193	12	BF688809	BF688809 IL3-ET011	C 326	28	2.0	420	9	AL386726	AL386726 MRC36E01
254	28	2.0	193	10	AM079315	AM079315 xB53a03.x	C 327	28	2.0	420	10	AM081116	AM081116 t87c06.x
255	28	2.0	198	9	A1703028	A1703028 w66h08.x	C 328	28	2.0	421	9	A1312829	A1312829 q774d07.x
256	28	2.0	199	10	AM264716	AM264716 xq34c07.x	C 329	28	2.0	422	13	BM005547	BM005547 Csc30218
257	28	2.0	202	9	A1473288	A1473288 t83b03.x	C 330	28	2.0	426	9	A1612042	A1612042 t63f01.x
258	28	2.0	202	9	A1700148	A1700148 t421a12.x	C 331	28	2.0	427	10	AM232309	AM232309 f117e07.x
259	28	2.0	209	10	BE87661	BE87661 176722.BA	C 332	28	2.0	430	9	A1802810	A1802810 t57b07.x
260	28	2.0	211	9	AA397361	AA397361 nc64a10.s	C 333	28	2.0	431	9	A1129332	A1129332 q48c05.x
261	28	2.0	214	9	A1783499	A1783499 t236d06.x	C 334	28	2.0	434	9	AA71560	AA71560 n205a03.s
262	28	2.0	220	9	A1452205	A1452205 mp85d08.x	C 335	28	2.0	435	10	BE510788	BE510788 946055E03
263	28	2.0	229	9	AU053807	AU053807 AU053807	C 336	28	2.0	441	13	B1278340	B1278340 UI-R-CW0
264	28	2.0	230	17	A2860301	A2860301 2M0166D08	C 337	28	2.0	448	9	A1942985	A1942985 t87c06.x
265	28	2.0	230	12	BF765579	BF765579 RC6-CS007	C 338	28	2.0	449	9	AU176517	AU176517 t87c06.x
266	28	2.0	238	10	BA494797	BA494797 BA494797	C 339	28	2.0	451	13	BM036376	BM036376 t87c06.x
267	28	2.0	238	10	BA494797	BA494797 BA494797	C 340	28	2.0	452	9	A1626591	A1626591 t87c06.x
268	28	2.0	238	14	N20169	N20169 yx41d02.s1	C 341	28	2.0	452	10	BE103182	BE103182 UI-R-BT1
269	28	2.0	244	12	BG629021	BG629021 CC-estf1L	C 342	28	2.0	454	9	A1834385	A1834385 t87c06.x
270	28	2.0	245	9	A1926147	A1926147 w044d04.x	C 343	28	2.0	457	9	A1601557	A1601557 t87c06.x
271	28	2.0	245	9	AA512293	AA512293 v119h01.r	C 344	28	2.0	457	9	AU034234	AU034234 t87c06.x
272	28	2.0	258	9	A1497871	A1497871 t89d06.x	C 345	28	2.0	464	13	BM381642	BM381642 t87c06.x
273	28	2.0	261	9	A1566460	A1566460 t86d05.x	C 346	28	2.0	464	9	A166418	A166418 t87c06.x
274	28	2.0	261	9	A1566460	A1566460 t86d05.x	C 347	28	2.0	464	13	BM381642	BM381642 t87c06.x
275	28	2.0	261	13	B1053062	B1053062 RC0-GN027	C 348	28	2.0	465	9	A1968158	A1968158 t87c06.x
276	28	2.0	262	9	A1245391	A1245391 qk23h03.x	C 349	28	2.0	468	9	A1221078	A1221078 t87c06.x
277	28	2.0	262	12	BG629401	BG629401 CC-estf1L	C 350	28	2.0	469	9	AU499310	AU499310 t87c06.x
278	28	2.0	266	12	BG627926	BG627926 CC-estf1L	C 351	28	2.0	469	10	BE207064	BE207064 t87c06.x
279	28	2.0	271	12	BF743901	BF743901 FM3-BT083	C 352	28	2.0	469	10	BE207064	BE207064 t87c06.x
280	28	2.0	279	12	BG630170	BG630170 CC-estf1L	C 353	28	2.0	471	14	BQ412857	BQ412857 t87c06.x
281	28	2.0	283	12	BG691756	BG691756 341066.BA	C 354	28	2.0	472	13	B1274076	B1274076 UI-R-CW0
282	28	2.0	291	12	BG691750	BG691750 341055.BA	C 355	28	2.0	474	10	AM212557	AM212557 t87c06.x
283	28	2.0	295	14	BM685008	BM685008 UI-R-EJ1-	C 356	28	2.0	478	10	AM129599	AM129599 t87c06.x
284	28	2.0	300	9	A1189405	A1189405 q001h02.x	C 357	28	2.0	478	12	BF415067	BF415067 t87c06.x
285	28	2.0	304	14	BQ265956	BQ265956 NISC_f110	C 358	28	2.0	480	14	BQ391048	BQ391048 t87c06.x
286	28	2.0	306	9	A1464411	A1464411 mw75d10.x	C 359	28	2.0	481	17	CNS00CXM	CNS00CXM t87c06.x
287	28	2.0	309	14	B0800158	B0800158 EST_7193	C 360	28	2.0	489	14	BQ392581	BQ392581 t87c06.x
288	28	2.0	312	12	BG628027	BG628027 CC-estf1L	C 361	28	2.0	491	9	A1942827	A1942827 t87c06.x
289	28	2.0	316	10	BE684373	BE684373 184756.MA	C 362	28	2.0	491	12	BF321705	BF321705 t87c06.x
290	28	2.0	317	9	A1453456	A1453456 t128b03.x	C 363	28	2.0	492	13	BM379828	BM379828 t87c06.x
291	28	2.0	319	9	A1252225	A1252225 q178f11.x	C 364	28	2.0	495	10	BE591255	BE591255 t87c06.x
292	28	2.0	332	9	A1674779	A1674779 w420c11.x	C 365	28	2.0	496	9	A1179433	A1179433 t87c06.x
293	28	2.0	335	9	AA829402	AA829402 oc06d07.s	C 366	28	2.0	500	10	AM554069	AM554069 t87c06.x
294	28	2.0	335	9	A1204157	A1204157 q370d04.x	C 367	28	2.0	503	9	AL0442238	AL0442238 t87c06.x
295	28	2.0	335	14	BQ161114	BQ161114 WHE0362.F	C 368	28	2.0	506	12	BQ444413	BQ444413 t87c06.x
296	28	2.0	338	14	BQ171201	BQ171201 WHE2339.G	C 369	28	2.0	506	14	BQ389223	BQ389223 t87c06.x
297	28	2.0	331	9	A1284218	A1284218 qj70a08.x	C 370	28	2.0	509	13	BM496766	BM496766 t87c06.x
298	28	2.0	332	13	B1416784	B1416784 hasp002xe	C 371	28	2.0	513	9	AU034515	AU034515 t87c06.x

C 372	28	2.0	513	14	BO393675	BO393675 NISC ng04	C 445	28	2.0	916	13	BM358187	BM358187 GA_Ea000
C 373	28	2.0	513	17	AZ081520	AZ081520 PL01A13T	C 446	28	2.0	920	14	BO706875	BO706875 AGENCOURT
C 374	28	2.0	523	14	BO640550	BO640550 he30b06.Y	C 447	28	2.0	924	14	BO708312	BO708312 AGENCOURT
C 375	28	2.0	524	12	BG378391	BG378391 UI-R-CUO-	C 448	28	2.0	938	12	BE789640	BE789640 601481591
C 376	28	2.0	525	9	A1184800	A1184800 qd24902.x	C 449	28	2.0	958	14	BO710664	BO710664 AGENCOURT
C 377	28	2.0	526	9	AA859629	AA859629 UI-R-E0-b	C 450	28	2.0	965	14	BO481781	BO481781 PV_GEA010
C 378	28	2.0	528	12	BF123286	BF123286 601759573	C 451	28	2.0	971	12	BF038942	BF038942 601461770
C 379	28	2.0	531	9	A1588466	A1588466 fc02e02.x	C 452	28	2.0	971	14	BO710803	BO710803 AGENCOURT
C 380	28	2.0	534	9	A1881387	A1881387 606067D06	C 453	28	2.0	1077	10	BE040152	BE040152 OD103C01
C 381	28	2.0	534	14	BO166735	BO166735 WHE0925.A	C 454	28	2.0	1080	13	BM477855	BM477855 AGENCOURT
C 382	28	2.0	535	14	BO800173	BO800173 EST_7208	C 455	28	2.0	1084	13	BM423738	BM423738 AGENCOURT
C 383	28	2.0	539	14	BO171235	BO171235 WHE2342.C	C 456	28	2.0	1103	12	BF578999	BF578999 602096115
C 384	28	2.0	541	14	BO520137	BO520137 NISC nl03	C 457	28	2.0	1110	13	BM455289	BM455289 AGENCOURT
C 385	28	2.0	543	12	BF179217	BF179217 601807468	C 458	28	2.0	1390	12	BG696442	BG696442 602659550
C 386	28	2.0	549	13	BM035015	BM035015 fu15b07.x	C 459	28	2.0	1568	12	BE965367	BE965367 601659280
C 387	28	2.0	551	13	BI698619	BI698619 ts030 the	C 460	28	2.0	2780	11	BC034257	BC034257 Mus muscu
C 388	28	2.0	554	12	BE876887	BE876887 601484785	C 461	27	1.9	71	9	AA878808	AA878808 6186102.s
C 389	28	2.0	558	14	BO746572	BO746572 UI-M-FAO-	C 462	27	1.9	73	14	H10854	H10854 ymo6f06.s1
C 390	28	2.0	559	17	BH098770	BH098770 RPCI-24-2	C 463	27	1.9	103	9	AA043685	AA043685 2K50D12.s
C 391	28	2.0	563	13	BI274368	BI274368 UI-R-CWO-	C 464	27	1.9	114	9	A1340731	A1340731 tb32b05.x
C 392	28	2.0	563	14	BO171616	BO171616 WHE1651.1	C 465	27	1.9	126	12	BF815930	BF815930 MR2-C1018
C 393	28	2.0	568	14	BO412099	BO412099 GA_Ed005	C 466	27	1.9	130	9	A1344802	A1344802 ED10606.x
C 394	28	2.0	569	14	BO403115	BO403115 GA_Ed005	C 467	27	1.9	130	12	BG834909	BG834909 353288 MA
C 395	28	2.0	572	14	BO524075	BO524075 NISC nc02	C 468	27	1.9	131	9	A1721723	A1721723 EC31F01.x
C 396	28	2.0	582	9	AJ501088	AJ501088 AJS01088	C 469	27	1.9	133	13	BI057403	BI057403 PM3-GN049
C 397	28	2.0	583	13	BM832282	BM832282 MEST547-B	C 470	27	1.9	137	12	BF811802	BF811802 MR2-C1018
C 398	28	2.0	588	9	AJ436044	AJ436044 AJ436044	C 471	27	1.9	138	13	BM196167	BM196167 C0323H12-
C 399	28	2.0	588	11	AY108823	AY108823 Zea mays	C 472	27	1.9	146	9	A1321686	A1321686 e2a07hm.f
C 400	28	2.0	589	14	BO906022	BO906022 L007G02.O	C 473	27	1.9	155	13	BM091655	BM091655 saq99h06.
C 401	28	2.0	594	10	BE439656	BE439656 HTM1-323F	C 474	27	1.9	159	10	AM105588	AM105588 xcd34906.x
C 402	28	2.0	598	14	BO796645	BO796645 EST_5583	C 475	27	1.9	159	13	BM091905	BM091905
C 403	28	2.0	600	12	BG874232	BG874232 MEST7-g1	C 476	27	1.9	160	9	A1587432	A1587432 c550f01.x
C 404	28	2.0	600	13	BI81321	BI81321 fu51c08.x	C 477	27	1.9	161	9	AJ381156	AJ381156 MEBB57C04
C 405	28	2.0	604	9	BO391602	BO391602 NISC mg19	C 478	27	1.9	161	9	AL123791	AL123791 AL723791
C 406	28	2.0	604	9	A1691770	A1691770 606008E05	C 479	27	1.9	161	10	BE138961	BE138961 c63f677.x
C 407	28	2.0	608	12	BF687533	BF687533 602066637	C 480	27	1.9	162	12	BF946591	BF946591 MR0-NN019
C 408	28	2.0	608	17	BH053362	BH053362 RPCI-24-3	C 481	27	1.9	167	12	BF245783	BF245783 601864279
C 409	28	2.0	611	10	AV763492	AV763492 AV763492	C 482	27	1.9	172	14	BO386509	BO386509 NISC ml18
C 410	28	2.0	613	2	HSM00482	HSM00482	C 483	27	1.9	175	9	A1345718	A1345718 CB284601.x
C 411	28	2.0	614	9	A1690947	A1690947 tq02e10.x	C 484	27	1.9	175	12	BE290887	BE290887 602387046
C 412	28	2.0	617	10	BE583519	BE583519 7-11B-HA	C 485	27	1.9	177	13	BM124689	BM124689 L0534F11-
C 413	28	2.0	619	13	BM073234	BM073234 MEST63-A1	C 486	27	1.9	184	14	BI916275	BI916275 603178184
C 414	28	2.0	621	10	AM968980	AM968980 EST81057	C 487	27	1.9	185	12	BF811784	BF811784 MR2-C1018
C 415	28	2.0	626	14	BO413419	BO413419 GA_Ed007	C 488	27	1.9	186	10	AM480006	AM480006 29125 MAR
C 416	28	2.0	627	14	BO415646	BO415646 GA_Ed010	C 489	27	1.9	199	13	BI849640	BI849640 477105 MA
C 417	28	2.0	637	9	A1231846	A1231846 EST228534	C 490	27	1.9	201	9	AJ499622	AJ499622 AJ499622
C 418	28	2.0	639	14	BO409937	BO409937 GA_Ed002	C 491	27	1.9	201	9	AJ500051	AJ500051 AJ500051
C 419	28	2.0	644	14	BO410459	BO410459 GA_Ed003	C 492	27	1.9	204	9	A1919310	A1919310 tuj32e11.x
C 420	28	2.0	654	11	AY068553	AY068553 Schmidtrea	C 493	27	1.9	204	14	BO612379	BO612379 sap59c02.
C 421	28	2.0	655	14	BO410346	BO410346 GA_Ed003	C 494	27	1.9	207	9	A1223676	A1223676 GX286c02.x
C 422	28	2.0	658	14	BM933376	BM933376 UI-M-BG2-	C 495	27	1.9	208	9	A1433992	A1433992 CB87b10.x
C 423	28	2.0	673	13	BI141680	BI141680 hasp001xn	C 496	27	1.9	208	9	A1344077	A1344077 CB87f10.x
C 424	28	2.0	679	9	AU057955	AU057955 AB0101A	C 497	27	1.9	208	9	A1610195	A1610195 wB80a04.x
C 425	28	2.0	687	10	BE038545	BE038545 AB0101A	C 498	27	1.9	210	13	BM023495	BM023495 iea81b12.y
C 426	28	2.0	702	10	AV715426	AV715426 AV715426	C 499	27	1.9	211	9	A1659962	A1659962 tC10c07.x
C 427	28	2.0	703	9	A1385746	A1385746 m421f06.x	C 500	27	1.9	211	10	AM023125	AM023125 d448f01.y
C 428	28	2.0	703	9	AL583616	AL583616 AL583616	C 501	27	1.9	213	12	BC691751	BC691751 341056 BA
C 429	28	2.0	717	10	AV760439	AV760439 AV760439	C 502	27	1.9	213	17	BH175041	BH175041 UMC-27 Ho
C 430	28	2.0	730	12	BF691895	BF691895 602248483	C 503	27	1.9	214	13	AU033864	AU033864 MR2-C1018
C 431	28	2.0	733	10	AV714894	AV714894 AV714894	C 504	27	1.9	214	12	BF814527	BF814527 MR2-C1018
C 432	28	2.0	739	14	BO747272	BO747272 UI-M-FAO-	C 505	27	1.9	216	10	AM302687	AM302687 x61c02.x
C 433	28	2.0	747	17	BH455427	BH455427 BOGHC31R	C 506	27	1.9	220	9	A1560101	A1560101 cp12g03.x
C 434	28	2.0	757	13	BM841131	BM841131 UI-R-DZO-	C 507	27	1.9	225	10	AMS56639	AMS56639 L0271D04-
C 435	28	2.0	769	12	BG620262	BG620262 602618546	C 508	27	1.9	225	9	A1391451	A1391451 tf96f03.x
C 436	28	2.0	772	13	BI888790	BI888790 ZF637-2.0	C 509	27	1.9	230	10	AM103363	AM103363 xc69h12.x
C 437	28	2.0	776	17	CNS05DUY	CNS05DUY	C 510	27	1.9	234	14	BO177041	BO177041 UI-M-DJ2-
C 438	28	2.0	776	17	AM009592	AM009592 ws84904.x	C 511	27	1.9	237	10	BM935892	BM935892 QV4-BT040
C 439	28	2.0	780	10	BF136590	BF136590 601779639	C 512	27	1.9	237	14	BE935892	BE935892 UI-M-BZ1-
C 440	28	2.0	781	17	BH420557	BH420557 BOH1135TR	C 513	27	1.9	239	10	BB056764	BB056764 BB056764
C 441	28	2.0	811	9	AJ301582	AJ301582 AJ301582	C 514	27	1.9	240	12	BE879397	BE879397 601484659
C 442	28	2.0	811	9	AJ301588	AJ301588 AJ301588	C 515	27	1.9	243	13	BM270602	BM270602 sak15c07.
C 443	28	2.0	848	12	BF140394	BF140394 601787710	C 516	27	1.9	244	10	BE233435	BE233435 139616 MA
C 444	28	2.0	848	12	BF140394	BF140394 601787710	C 517	27	1.9	245	10	AM046394	AM046394 UI-M-BH1-

C 518	27	1.9	245	12	BG115495	BG115495 602316651	591	27	1.9	356	12	BG115428	BG115428 602284855
C 519	27	1.9	246	9	AI003423	AI003423 AK87B07.8	C 592	27	1.9	357	14	BQ168481	BQ168481 WHE2112.D
C 520	27	1.9	247	9	AI766607	AI766607 W101D09.x	C 593	27	1.9	358	14	BQ915781	BQ915781 CHB15M02
C 521	27	1.9	247	10	AM118552	AM118552 xer9e11.x	C 594	27	1.9	359	9	AI926661	AI926661 w049d04.x
C 522	27	1.9	250	9	AI036951	AI036951 wh23d03.x	C 595	27	1.9	362	10	BE664946	BE664946 152857.MA
C 523	27	1.9	257	9	AL631587	AL631587 AL631587	C 596	27	1.9	362	12	BF428026	BF428026 daa01a11.
C 524	27	1.9	260	13	BM074806	BM074806 MEST297-H	C 597	27	1.9	363	9	AI473843	AI473843 ca17f02.x
C 525	27	1.9	261	10	AM316339	AM316339 14677.MAR	C 598	27	1.9	364	9	AI054351	AI054351 q164a10.x
C 526	27	1.9	263	9	AL371879	AL371879 MEB47A12	C 599	27	1.9	366	12	BG034511	BG034511 602302629
C 527	27	1.9	264	17	AZ826848	AZ826848 2M0102E21	C 600	27	1.9	366	14	C97562	C97562 C97562.Rice
C 528	27	1.9	265	10	AM214086	AM214086 uc03b09.x	C 601	27	1.9	367	12	BF426061	BF426061 gs05a03.Y
C 529	27	1.9	265	14	BQ392389	BQ392389 NISC.mq24	C 602	27	1.9	368	9	AA287771	AA287771 nc14q11.S
C 530	27	1.9	269	9	AI591260	AI591260 lf75C09.x	C 603	27	1.9	368	12	BE886995	BE886995 601506478
C 531	27	1.9	270	14	BQ612980	BQ612980 sap79e04.x	C 604	27	1.9	369	9	AI650796	AI650796 wa19g05.x
C 532	27	1.9	270	14	BQ418129	BQ418129 iK52d03.x	C 605	27	1.9	370	10	BB808982	BB808982 BB808982
C 533	27	1.9	273	10	AM317330	AM317330 6G47G05.Y	C 606	27	1.9	373	10	BQ400888	BQ400888 NISC.m011
C 534	27	1.9	274	9	AI345144	AI345144 ts85h11.x	C 607	27	1.9	373	9	AL372321	AL372321 MEB4A0B06
C 535	27	1.9	276	9	AI335371	AI335371 lb79D03.x	C 608	27	1.9	374	9	AI552215	AI552215 v168d11.Y
C 536	27	1.9	276	9	AI152617	AI152617 mq45a01.x	C 609	27	1.9	374	13	BI341595	BI341595 369072.MA
C 537	27	1.9	277	12	BF766169	BF766169 IL2-CS005	C 610	27	1.9	375	9	AI266586	AI266586 qn29a12.x
C 538	27	1.9	278	10	AM022306	AM022306 df36f05.Y	C 611	27	1.9	375	9	AI633184	AI633184 tz07C03.x
C 539	27	1.9	278	14	C96755	C96755 C96755.Rice	C 612	27	1.9	376	12	BG735006	BG735006 CC-eflCL
C 540	27	1.9	280	14	BQ168535	BQ168535 WHE2119_E	C 613	27	1.9	376	13	BI492265	BI492265 df21h03.w
C 541	27	1.9	286	9	AI470305	AI470305 lf42a01.x	C 614	27	1.9	376	14	BM703213	BM703213 UI-E-CL1-
C 542	27	1.9	287	9	AI914738	AI914738 lf01d07.x	C 615	27	1.9	378	9	AI765745	AI765745 w184a05.x
C 543	27	1.9	287	9	AA260997	AA260997 vD04h03.x	C 616	27	1.9	379	14	BQ384813	BQ384813 NISC.m008
C 544	27	1.9	288	9	AL714200	AL714200 AL714200	C 617	27	1.9	382	9	AU093432	AU093432 AU093432
C 545	27	1.9	289	9	AL642390	AL642390 AL642390	C 618	27	1.9	382	9	AI068399	AI068399 WHE2104.C
C 546	27	1.9	290	9	AI766627	AI766627 w101f09.x	C 619	27	1.9	382	14	BQ168392	BQ168392 CC-eflCL
C 547	27	1.9	291	14	BM891646	BM891646 sam42c01.	C 620	27	1.9	383	12	BG734918	BG734918 CC-eflCL
C 548	27	1.9	293	9	AI245538	AI245538 qK25d12.x	C 621	27	1.9	384	17	AZ555448	AZ555448 mY30E11.Y
C 549	27	1.9	294	14	R76691	R76691 y161f01.s1	C 622	27	1.9	384	17	AA255448	AA255448 RPCL-23-2
C 550	27	1.9	295	14	BM696042	BM696042 UI-E-CL1-	C 623	27	1.9	385	13	BG941932	BG941932 ax19b01.x
C 551	27	1.9	296	14	AL372070	AL372070 MEB48D06	C 624	27	1.9	387	9	AI128115	AI128115 q333d04.x
C 552	27	1.9	296	14	BM878753	BM878753 UI-H-COO-	C 625	27	1.9	388	14	BM874824	BM874824 lf36C05.x
C 553	27	1.9	297	9	AI342627	AI342627 qC33f02.x	C 626	27	1.9	389	12	BG794320	BG794320 UTSW.SM20
C 554	27	1.9	298	9	AI275177	AI275177 qW08d08.x	C 627	27	1.9	392	9	AI538881	AI538881 lf74h05.x
C 555	27	1.9	298	9	AI395789	AI395789 MA005149.	C 628	27	1.9	392	9	AL640266	AL640266 AL640266
C 556	27	1.9	298	9	AL377895	AL377895 MEBB34F07	C 629	27	1.9	393	9	AL640879	AL640879 AL640879
C 557	27	1.9	298	9	AU184667	AU184667 AU184667	C 630	27	1.9	394	9	AI376638	AI376638 MEBB25B06
C 558	27	1.9	311	9	AI702511	AI702511 tz66h03.x	C 631	27	1.9	396	9	AI636954	AI636954 ts86a07.x
C 559	27	1.9	312	12	BG511817	BG511817 sad1ld07.	C 632	27	1.9	396	14	BQ538550	BQ538550 MEST602-D
C 560	27	1.9	313	9	AI559498	AI559498 lf44f10.x	C 633	27	1.9	397	14	BQ792662	BQ792662 EST.8382
C 561	27	1.9	313	9	AI864823	AI864823 wQ4G12.x	C 634	27	1.9	397	9	AI185391	AI185391 qe52d10.x
C 562	27	1.9	313	10	AM394528	AM394528 sh33b07.Y	C 635	27	1.9	397	14	BM921005	BM921005 AGENCOURT
C 563	27	1.9	320	17	AO973548	AO973548 RPCL-23-3	C 636	27	1.9	400	10	AV736294	AV736294 AV736294
C 564	27	1.9	322	12	BG652947	BG652947 sad43a05.	C 637	27	1.9	401	9	AI331854	AI331854 lf26a04.x
C 565	27	1.9	327	9	AL648908	AL648908 AL648908	C 638	27	1.9	401	13	BM201248	BM201248 C0222B11-
C 566	27	1.9	327	12	BE723483	BE723483 193314.MA	C 639	27	1.9	402	9	AL649148	AL649148 AL649148
C 567	27	1.9	328	10	AV743641	AV743641 AV743641	C 640	27	1.9	402	12	BG628532	BG628532 CC-eflCL
C 568	27	1.9	329	17	BH341290	BH341290 CH230-183	C 641	27	1.9	402	12	BR343283	BR343283 602015943
C 569	27	1.9	330	9	AA063665	AA063665 ESTM239F	C 642	27	1.9	403	9	AI274542	AI274542 qY10f07.x
C 570	27	1.9	330	9	AI241258	AI241258 qK16a09.x	C 643	27	1.9	403	12	BG899928	BG899928 HOA40-1-D
C 571	27	1.9	332	12	BG631246	BG631246 CC-eflCL	C 644	27	1.9	404	9	AI686800	AI686800 lf79d08.x
C 572	27	1.9	332	12	BG670910	BG670910 DRNB1D12	C 645	27	1.9	407	9	AL637818	AL637818 AL637818
C 573	27	1.9	332	14	BQ524538	BQ524538 NISC.no05	C 646	27	1.9	408	9	AU033082	AU033082 AU033082
C 574	27	1.9	333	9	AL657667	AL657667 AL657667	C 647	27	1.9	409	10	BR470826	BR470826 WHE0281.C
C 575	27	1.9	333	14	BQ771385	BQ771385 UI-M-F10-	C 648	27	1.9	410	9	AI167815	AI167815 o290d12.x
C 576	27	1.9	334	12	BG629042	BG629042 CC-eflCL	C 649	27	1.9	411	12	BG178669	BG178669 602328553
C 577	27	1.9	337	9	AI239962	AI239962 qH4d11.x	C 650	27	1.9	413	14	BM877594	BM877594 lf90b06.x
C 578	27	1.9	337	9	AI590833	AI590833 tw88c02.x	C 651	27	1.9	413	9	AI428560	AI428560 m043609.x
C 579	27	1.9	340	9	BO794136	BO794136 EST.3074	C 652	27	1.9	414	9	AI863335	AI863335 tz39c02.x
C 580	27	1.9	341	9	AI430711	AI430711 mc61f04.Y	C 653	27	1.9	415	12	BG181015	BG181015 602329114
C 581	27	1.9	343	9	AI661477	AI661477 va36h12.x	C 654	27	1.9	415	14	BQ520473	BQ520473 NISC.m104
C 582	27	1.9	344	10	AM406745	AM406745 UI-HF-B10	C 655	27	1.9	415	17	BQ800244	BQ800244 EST.7279
C 583	27	1.9	346	9	AI571960	AI571960 lf71g02.x	C 656	27	1.9	417	10	BE404569	BE404569 WHE0443.H
C 584	27	1.9	349	12	BF198644	BF198644 248556.MA	C 657	27	1.9	420	10	BE667455	BE667455 151822.MA
C 585	27	1.9	350	9	AM597372	AM597372 a192a02.Y	C 658	27	1.9	421	10	AM074476	AM074476 xa94c11.x
C 586	27	1.9	351	9	AL509243	AL509243 AL509243	C 659	27	1.9	421	14	BQ388394	BQ388394 NISC.mq01
C 587	27	1.9	351	9	AA687965	AA687965 nv13h09.8	C 660	27	1.9	422	9	AI683519	AI683519 cx67e01.x
C 588	27	1.9	354	9	AA185574	AA185574 tm49f11.r	C 661	27	1.9	422	10	BE032048	BE032048 130853.MA
C 589	27	1.9	355	9	AI560546	AI560546 lf48g08.x	C 662	27	1.9	422	10	BE032048	BE032048 kh43d11.Y
C 590	27	1.9	355	9	AI560546	AI560546 lf48g08.x	C 663	27	1.9	423	13	BI782848	BI782848

C 810	27	1.9	587	13	BM026567	ft07d04.x	883	27	1.9	728	12	8780595	8780595 SEAMC000
C 811	27	1.9	589	14	BM074675	UI-M-FAO-	C 884	27	1.9	730	17	B18656	B18656 F2IC2-T7 IG
C 812	27	1.9	590	14	BM086843	NISC_mn20	C 885	27	1.9	736	12	BF839398	BF839398 602045517
C 813	27	1.9	590	14	BM051464	NISC_n111	C 886	27	1.9	738	10	AM667768	AM667768 EA Ea001
C 814	27	1.9	591	14	BM26821	W26821 13b4 Human	C 887	27	1.9	738	13	B1455733	B1455733 603171738
C 815	27	1.9	592	14	BM033133	E238401.Y	C 888	27	1.9	739	10	B6119500	B6119500 601473053
C 816	27	1.9	592	14	BM283548	A2283548 RPCI-23-1	C 889	27	1.9	742	10	AV757936	AV757936 AV757936
C 817	27	1.9	593	14	BM074573	UI-M-FAO-	C 890	27	1.9	743	10	AV702718	AV702718 AV702718
C 818	27	1.9	593	14	BM1438540	UI-M-FAO-	C 891	27	1.9	743	10	AV757929	AV757929 AV757929
C 819	27	1.9	593	14	BM074573	UI-M-FAO-	C 892	27	1.9	743	10	AV757929	AV757929 AV757929
C 820	27	1.9	599	10	BM074573	UI-M-FAO-	C 893	27	1.9	747	14	BM074677	BM074677 UI-M-FAO-
C 821	27	1.9	608	10	BM074573	UI-M-FAO-	C 894	27	1.9	750	9	BM074677	BM074677 UI-M-FAO-
C 822	27	1.9	608	14	BM074573	UI-M-FAO-	C 895	27	1.9	751	10	BM074677	BM074677 UI-M-FAO-
C 823	27	1.9	612	14	BM0391527	NISC_mq19	C 896	27	1.9	751	17	B21660	B21660 F10L20-Sp6
C 824	27	1.9	613	13	BM080627	BM080627 NEST522-F	C 897	27	1.9	755	12	BF008661	BF008661 MY1100160
C 825	27	1.9	614	14	BM0394809	NISC_mq11	C 898	27	1.9	755	12	BM0397364	BM0397364 602439149
C 826	27	1.9	623	13	BM271848	BM271848 I935T08.Y	C 899	27	1.9	763	12	BM265448	BM265448 HV-EA001
C 827	27	1.9	626	9	AL589615	AL589615 DXF2P451L	C 900	27	1.9	765	12	BM0894028	BM0894028 601437841
C 828	27	1.9	626	14	BM089385	BM089385 NISC_mq07	C 901	27	1.9	766	9	AI098125	AI098125 ue32c10.x
C 829	27	1.9	627	10	BM0327777	BM0327777 dr01c11.Y	C 902	27	1.9	771	12	BM086522	BM086522 H3137C06-
C 830	27	1.9	630	12	BM0878434	BM0878434 601487833	C 903	27	1.9	786	17	BM19233	BM19233 F16B8177 IG
C 831	27	1.9	630	17	BM2074	F2074-F28C23-T7 I	C 904	27	1.9	787	10	BM117147	BM117147 f63f601.x
C 832	27	1.9	631	14	BM0390775	NISC_mq15	C 905	27	1.9	795	12	BM08630	BM08630 MY1100157
C 833	27	1.9	631	14	BM0390775	NISC_mq15	C 906	27	1.9	802	12	BM159370	BM159370 601770348
C 834	27	1.9	632	9	AL514517	AL514517 NISC_mq05	C 907	27	1.9	804	11	AV107483	AV107483 Zeal maye
C 835	27	1.9	632	10	BM032858	BM032858 AV732858	C 908	27	1.9	805	9	AL515157	AL515157 ENTP185TR
C 836	27	1.9	632	10	BM049073	BM049073 xf09f03.x	C 909	27	1.9	815	17	BM074677	BM074677 ENTP185TR
C 837	27	1.9	632	17	BM0502774	BM0502774 1M0342815	C 910	27	1.9	817	13	BM1827006	BM1827006 603075530
C 838	27	1.9	637	11	BM019228	BM019228 Homo BABI	C 911	27	1.9	819	17	BM0900251	BM0900251 HS-1129.A
C 839	27	1.9	638	11	BM0288624	BM0288624 602385407	C 912	27	1.9	821	12	BM038309	BM038309 601458133
C 840	27	1.9	641	12	BM0411214	BM0411214 GA1-E0003	C 913	27	1.9	822	12	BM0876977	BM0876977 601484744
C 841	27	1.9	641	12	BM0899722	BM0899722 HOA39-1-F	C 914	27	1.9	831	9	AA989771	AA989771 ue16h08.x
C 842	27	1.9	645	14	BM0413604	BM0413604 EA Ed007	C 915	27	1.9	836	13	BM1888418	BM1888418 ZF637-2-0
C 843	27	1.9	649	9	AA753270	AA753270 96BS0301	C 916	27	1.9	837	12	BM086836	BM086836 602112129
C 844	27	1.9	649	9	AL662262	AL662262 AL662262	C 917	27	1.9	840	12	BM087096	BM087096 602102901
C 845	27	1.9	649	13	BM145667	BM145667 602911308	C 918	27	1.9	845	13	BM0976714	BM0976714 602847054
C 846	27	1.9	655	10	BM0628793	BM0628793 BM628793	C 919	27	1.9	852	12	BM0536401	BM0536401 602051760
C 847	27	1.9	655	13	BM014072	BM014072 603639733	C 920	27	1.9	868	9	AL514007	AL514007 602051760
C 848	27	1.9	658	10	BM067756	BM067756 EA Ea001	C 921	27	1.9	869	12	BM0867300	BM0867300 602102342
C 849	27	1.9	663	9	AL514595	AL514595 AL514595	C 922	27	1.9	875	9	AL515001	AL515001 AL515001
C 850	27	1.9	664	13	BM080007	BM080007 MEST102-F	C 923	27	1.9	875	10	BM0758496	BM0758496 AV758496
C 851	27	1.9	668	14	BM0748094	BM0748094 UI-M-FAO-	C 924	27	1.9	879	14	BM0430932	BM0430932 AGENCOURT
C 852	27	1.9	671	17	BM0570870	BM0570870 281PVB01	C 925	27	1.9	882	12	BM0768345	BM0768345 602084970
C 853	27	1.9	672	17	BM0569161	BM0569161 256PVB03	C 926	27	1.9	883	13	BM119916	BM119916 L0932D04-
C 854	27	1.9	680	9	AL717941	AL717941 AL717941	C 927	27	1.9	890	13	BM0765031	BM0765031 603051174
C 855	27	1.9	680	12	BM032214	BM032214 602301968	C 928	27	1.9	891	13	BM0962643	BM0962643 AGENCOURT
C 856	27	1.9	681	9	AL513595	AL513595 AL513595	C 929	27	1.9	898	12	BM0681161	BM0681161 602628994
C 857	27	1.9	683	17	BM0855106	BM0855106 Pan tC091	C 930	27	1.9	900	13	BM0453906	BM0453906 AGENCOURT
C 858	27	1.9	683	13	BM085547	BM085547 5013c7 NI	C 931	27	1.9	906	14	BM0424492	BM0424492 AGENCOURT
C 859	27	1.9	684	13	BM1673633	BM1673633 fct29f11.x	C 932	27	1.9	909	12	BM0857813	BM0857813 601653946
C 860	27	1.9	685	9	AA722498	AA722498 2N31f04.8	C 933	27	1.9	912	12	BM0729852	BM0729852 602107805
C 861	27	1.9	688	10	BM0561959	BM0561959 601347174	C 934	27	1.9	918	14	BM0729852	BM0729852 AGENCOURT
C 862	27	1.9	689	9	AL736282	AL736282 6D26C08.Y	C 935	27	1.9	927	14	BM0726633	BM0726633 AGENCOURT
C 863	27	1.9	690	17	BM081003	BM081003 Pan tC091	C 936	27	1.9	929	9	AL514685	AL514685 AGENCOURT
C 864	27	1.9	693	12	BM0900037	BM0900037 HOA48-1-H	C 937	27	1.9	929	12	BM08289728	BM08289728 602385016
C 865	27	1.9	699	17	BM19539	BM19539 T2DD12-T7 I	C 938	27	1.9	932	11	BM0105961	BM0105961 Zea maye
C 866	27	1.9	702	12	BM0296669	BM0296669 602393612	C 939	27	1.9	932	13	BM1115400	BM1115400 602663259
C 867	27	1.9	703	10	BM0233560	BM0233560 BB623560	C 940	27	1.9	932	14	BM0220024	BM0220024 AGENCOURT
C 868	27	1.9	705	12	BM0281274	BM0281274 EST445781	C 941	27	1.9	933	12	BM0789680	BM0789680 601481449
C 869	27	1.9	709	9	BM056676	BM056676 602117503	C 942	27	1.9	936	14	BM0717383	BM0717383 AGENCOURT
C 870	27	1.9	709	9	BM056676	BM056676 602117503	C 943	27	1.9	942	9	AL515269	AL515269 AGENCOURT
C 871	27	1.9	713	12	BM063552	BM063552 603288008	C 944	27	1.9	945	12	BM0301994	BM0301994 6020302832
C 872	27	1.9	713	12	BM063552	BM063552 603288008	C 945	27	1.9	958	13	BM0736754	BM0736754 602360453
C 873	27	1.9	714	9	BM096693	BM096693 601592078	C 946	27	1.9	959	14	BM0425085	BM0425085 AGENCOURT
C 874	27	1.9	717	12	BM0688990	BM0688990 602320372	C 947	27	1.9	962	14	BM0618988	BM0618988 RMOSE02C1
C 875	27	1.9	717	12	BM0778635	BM0778635 601466185	C 948	27	1.9	962	14	BM0619158	BM0619158 RMOSE04D0
C 876	27	1.9	718	17	BM19528	BM19528 F2P232-T7 I	C 949	27	1.9	962	14	BM0619495	BM0619495 RMOSE08D0
C 877	27	1.9	720	12	BM0289378	BM0289378 602381405	C 950	27	1.9	962	14	BM0619516	BM0619516 RMOSE01B1
C 878	27	1.9	721	17	BM0502296	BM0502296 RPCI-24-3	C 951	27	1.9	962	14	BM0295642	BM0295642 AGENCOURT
C 879	27	1.9	721	17	BM0562203	BM0562203 BOHPY63TF	C 952	27	1.9	975	12	BM0253070	BM0253070 602365596
C 880	27	1.9	722	12	BM0345610	BM0345610 d489f07.Y	C 953	27	1.9	984	12	BM0781975	BM0781975 601470378
C 881	27	1.9	724	9	BM0529570	BM0529570 ui82c01.x	C 954	27	1.9	985	12	BM0874214	BM0874214 601484105
C 882	27	1.9	724	14	BM087385	BM087385 UI-H-DF0-	C 955	27	1.9	991	13	BM0910943	BM0910943 6020808728

956	27	1.9	984	12	BS635590	BS635590	602582319
957	27	1.9	998	13	BM416548	BM416548	OP21634 M
958	27	1.9	1001	14	BQ729018	BQ729018	AGENCOURT
959	27	1.9	1006	14	BQ877143	BQ877143	AGENCOURT
960	27	1.9	1010	12	BF797153	BF797153	602257824
961	27	1.9	1013	14	BQ727354	BQ727354	AGENCOURT
962	27	1.9	1014	14	BQ950220	BQ950220	AGENCOURT
963	27	1.9	1016	17	CNS03237	AL255280	Tetradon
964	27	1.9	1025	12	BF034237	BF034237	601456031
965	27	1.9	1027	14	BQ844400	BQ844400	AGENCOURT
966	27	1.9	1040	14	BQ944642	BQ944642	AGENCOURT
967	27	1.9	1049	13	BM553565	BM553565	AGENCOURT
968	27	1.9	1051	12	BF168559	BF168559	601775462
969	27	1.9	1056	9	AL513931	AL513931	AL513931
970	27	1.9	1060	12	BF033752	BF033752	601454087
971	27	1.9	1063	13	BT743239	BT743239	603351348
972	27	1.9	1069	13	BG912561	BG912561	602806674
973	27	1.9	1076	12	BE962688	BE962688	601656062
974	27	1.9	1082	14	BQ930141	BQ930141	AGENCOURT
975	27	1.9	1085	13	BM564621	BM564621	AGENCOURT
976	27	1.9	1088	10	BE621067	BE621067	601493674
977	27	1.9	1090	14	BM907065	BM907065	AGENCOURT
978	27	1.9	1103	12	BM418955	BM418955	602779368
979	27	1.9	1105	13	BM455446	BM455446	AGENCOURT
980	27	1.9	1107	14	BQ946165	BQ946165	AGENCOURT
981	27	1.9	1110	14	BM928352	BM928352	AGENCOURT
982	27	1.9	1155	10	BE420976	BE420976	HMM004.E0
983	27	1.9	1162	13	BM476725	BM476725	AGENCOURT
984	27	1.9	1191	11	AY104119	AY104119	Zea mays
985	27	1.9	1211	14	BQ929879	BQ929879	AGENCOURT
986	27	1.9	1216	13	BG035441	BG035441	602325071
987	27	1.9	1217	13	BI688046	BI688046	603314622
988	27	1.9	1321	12	BF179216	BF179216	601807467
989	27	1.9	1328	14	BM467957	BM467957	AGENCOURT
990	27	1.9	1361	14	BM926544	BM926544	AGENCOURT
991	27	1.9	1430	14	BM805984	BM805984	AGENCOURT
992	27	1.9	1468	13	BM545669	BM545669	AGENCOURT
993	27	1.9	1583	12	BG252814	BG252814	602365383
994	27	1.9	1771	11	BE778572	BE778572	601466103
995	27	1.9	1771	11	AY103565	AY103565	Zea mays
996	27	1.9	1807	12	BE889423	BE889423	601512540
997	27	1.9	2020	11	AY103619	AY103619	Zea mays
998	27	1.9	2100	9	AU183456	AU183456	ft239312.x
999	26	1.9	50	13	BI865491	BI865491	
1000	26	1.9	59	9	AL802723	AL802723	

ALIGNMENTS

RESULT 1	BM927766	1039 bp	mrna	linear	EST 12-MAR-2002			
LOCUS	AGENCOURT_6729770	NIH_MGC_100	Homo sapiens	CDNA clone	IMAGE:5797598			
DEFINITION	5', mRNA sequence.							
ACCESSION	BM927766	GI:19378145						
VERSION	EST.							
KEYWORDS	human.							
SOURCE	Homo sapiens							
ORGANISM	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.							
REFERENCE	1 (bases 1 to 1039)							
AUTHORS	NIH-MGC http://mgc.nci.nih.gov/ .							
TITLE	National Institutes of Health, Mammalian Gene Collection (MGC)							
JOURNAL	Unpublished (1999)							
COMMENT	Contact: Robert Strausberg, Ph.D. Email: rgs@bbs.fda.gov Tissue Procurement: CGAP (Stanford) CDNA Library Preparation: Rubin Laboratory CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNLN) DNA Sequencing by: Agencourt Bioscience Corporation Clone distribution: MGC clone distribution information can be							

FEATURES

Found through the I.M.A.G.E. Consortium/LNLN at:
http://image.llnl.gov
Plate: L10M2024 row: d column: 15
High quality sequence stop: 655.
Location/Qualifiers
1..1039

source

/organism="Homo sapiens"

/db_xref="taxon:9606"

/clone_lib="NIH_MGC_100"

/tissue_type="hepatocellular carcinoma, cell line"

/lab_host="DH10B (phage-resistant)"

/note="Organ: liver; Vector: pOTB7; Site: 1: XhoI; Site: 2: EcoRI; CDNA made by oligo-dt priming. Directionally cloned into EcoRI/XhoI sites using the following 5' adaptor: GGCACGAG(G). Size-selected 250bp for average insert size 1.8kb. Library constructed by Ling Hong in the laboratory of Gerald M. Rubin (University of California, Berkeley) using ZAP-CDNA synthesis kit (Stratagene) and Superscript II RT (Life Technologies). Note: this is a NIH_MGC Library."

BASE COUNT

223 a 276 c 300 g 237 t 3 others

ORIGIN

Query Match

Best Local Similarity 99.8%; Pred. No. 0;

Matches 657; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY	442	CAGGTGGAGCTGTCTTTTCAAGATGTGAGAGCCAGTGTCTCTGATCCCAACA	501
DB	94	CAGGTGGAGCTGTCTTTTCAAGATGTGAGAGCCAGTGTCTCTGATCCCAACA	153
QY	502	AAGACCATATGTGAAGGCTGCTGATCCATATGAGGGCTGGGCTGACAGTACT	561
DB	154	AAGACCATATGTGAAGGCTGCTGATCCATATGAGGGCTGGGCTGACAGTACT	213
QY	562	ATCCTGAGAGCTGGGCTTGCCTGAGAGGAGTGAATTCGACATGGCAGACATGTC	621
DB	214	ATCCTGAGAGCTGGGCTTGCCTGAGAGGAGTGAATTCGACATGGCAGACATGTC	273
QY	622	ACCTGGGAACCCCTGACGAAAGTATACATCCGACAGACATGTATACGACGAA	681
DB	274	ACCTGGGAACCCCTGACGAAAGTATACATCCGACAGACATGTATACGACGAA	333
QY	682	CGGCAATATATGCAATATTAATATGAGTTTACACCTAGTATGAGACTGTGGC	741
DB	334	CGGCAATATATGCAATATTAATATGAGTTTACACCTAGTATGAGACTGTGGC	393
QY	742	TCTTATCCAGAAATATGAGGATATGATGCTCTCCAACTGTGGCTGTAAGCAAG	801
DB	394	TCTTATCCAGAAATATGAGGATATGATGCTCTCCAACTGTGGCTGTAAGCAAG	453
QY	802	CTCAGGCTATGCTCCCACTGGGGGCTGTGCCCTCTCCGAGAGGTTCCGTGGAGCC	861
DB	454	CTCAGGCTATGCTCCCACTGGGGGCTGTGCCCTCTCCGAGAGGTTCCGTGGAGCC	513
QY	862	CCATCACTGTTCAATATGAGATGAGATGAGTAAAGCCCTGCTGCTGCTGACACA	921
DB	514	CCATCACTGTTCAATATGAGATGAGATGAGTAAAGCCCTGCTGCTGCTGACACA	573
QY	922	TGCAACAGAGGCGGTGGGGCTGCTGGGGCAATCCATTCGTGAGTGTCTTCAGCT	981
DB	574	TGCAACAGAGGCGGTGGGGCTGCTGGGGCAATCCATTCGTGAGTGTCTTCAGCT	633
QY	982	TAGGTGTGACAGAGACTTGGGGGAGTGTCCAGATGTGGGTGATTTCTTACTGG	1041
DB	634	TAGGTGTGACAGAGACTTGGGGGAGTGTCCAGATGTGGGTGATTTCTTACTGG	693
QY	1042	GGAGGCTATCTTGACCTCCGACAGGGGACACTCCCA	1079
DB	694	GGAGGCTATCTTGACCTCCGACAGGGGACACTCCCA	731

RESULT 2
BG827019 776 bp mRNA linear EST 22-MAY-2001
LOCUS 602749053F1 NIH_MGC_17 Homo sapiens cDNA clone IMAGE:4901869 5'
DEFINITION mRNA sequence.
ACCESSION BG827019
VERSION BG827019.1 GI:14174606
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
REFERENCE 1 (bases 1 to 776)
NIH-MGC <http://mgi.nci.nih.gov/>.
National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgabs-r@mail.nih.gov
Tissue Procurement: ATCC
CDNA Library Preparation: Ling Hong/Rubin Laboratory
DNA Sequencing by: Invitrogen Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LINL at:
<http://image.lnl.gov>
Plate: LICM1797 row: 3 column: 14
High quality sequence stop: 771.
Location/Qualifiers
1..776
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:4901869"
/clone_lib="NIH MGC 17"
/tissue_type="rhodomyosarcoma"
/lab_host="DH10B (phage-resistant)"
/note="Organ: muscle; Vector: pOTB7, Site_1: EcoRI;
Site_2: XhoI; cDNA made by oligo-dT priming.
Directionally cloned into EcoRI/XhoI sites using the
following 5' adaptor: GCGACGAG(G). Site-selected >500bp
for average insert size 1.8kb. Library constructed by
Ling Hong in the laboratory of Gerald M. Rubin (University
of California, Berkeley) using ZAP-CDNA synthesis kit
(Stratagene) and Superscript II RT (Life Technologies)."
BASE COUNT 158 a 228 c 222 g 168 t
ORIGIN
Query Match 40.5%; Score 566; DB 12; Length 776;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 566; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
442 CAGGTGGGAGGCTGGTCTTTTCAAGCTGGATGGAGCCAGTGGTCCCTGATCCCAACA 501
|||||
DB 202 CAGGTGGGAGGCTGGTCTTTTCAAGCTGGATGGAGCCAGTGGTCCCTGATCCCAACA 261
|||||
QY 502 AGACCAATATGTGAAGGCTCTGGCTGACCTATCTGAGGGCTCGCTGACCAAGCTGACT 561
|||||
DB 262 AGACCAATATGTGAAGGCTCTGGCTGACCTATCTGAGGGCTCGCTGACCAAGCTGACT 321
|||||
QY 562 ATCTCAGACAGCTGGCTTCCCTGTGGAGGAGTACTGCTGACCTGGACAGCTGATGTC 621
|||||
DB 322 ATCTCAGACAGCTGGCTTCCCTGTGGAGGAGTACTGCTGACCTGGACAGCTGATGTC 381
|||||
QY 622 ACCTGGGAACCCCTGGAGACAAAGCTAACATCCACAGACAGAGATGTGACAGGACAA 681
|||||
DB 382 ACCTGGGAACCCCTGGAGACAAAGCTAACATCCACAGACAGAGATGTGACAGGACAA 441
|||||
QY 682 CGTGCAATATGCAATATGTTAAATGTGATTACAGCTGATGGAGTGGCTGGC 741
|||||
DB 442 CGTGCAATATGCAATATGTTAAATGTGATTACAGCTGATGGAGTGGCTGGC 501
|||||
QY 742 TCTTAGTCCAGGAATATGAGGGGTATGATGCTCTTCCAACTGTGGGCTGTAAAGCAAG 801
|||||
DB 502 TCTTAGTCCAGGAATATGAGGGGTATGATGCTCTTCCAACTGTGGGCTGTAAAGCAAG 561
|||||

QY 802 CTCAGGCTAGTCTCCCACTGGGGGGCTGGTCCCTGCTGGAGACGTTCCGTGGCAGCC 861
|||||
DB 562 CTCAGGCTAGTCTCCCACTGGGGGGCTGGTCCCTGCTGGAGACGTTCCGTGGCAGCC 621
|||||
QY 862 CCATCACTGTGTCAATAGTGTAGAGATGACTTAAAGCCCTGCTGCTGCTGCACA 921
|||||
DB 622 CCATCACTGTGTCAATAGTGTAGAGATGACTTAAAGCCCTGCTGCTGCTGCACA 681
|||||
QY 922 TGGCAGACAGGCGGCTGGGGGCTGCTGGGGACATCCATCCGTGAGTTCCTAGCT 981
|||||
DB 682 TGGCAGACAGGCGGCTGGGGGCTGCTGGGGACATCCATCCGTGAGTTCCTAGCT 741
|||||
QY 982 TAGGTCTGACAGAGAGACTTGGCGGG 1007
|||||
DB 742 TAGGTCTGACAGAGAGACTTGGCGGG 767
|||||

RESULT 3
BQ013425/c 520 bp mRNA linear EST 26-MAR-2002
LOCUS UI-1-BC1p-ayx-c-01-0-UI s1 NCI CGAP_P13 Homo sapiens cDNA clone
DEFINITION UI-1-BC1p-ayx-c-01-0-UI 3', mRNA sequence.
ACCESSION BQ013425
VERSION BQ013425.1 GI:19738326
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
REFERENCE 1 (bases 1 to 520)
NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgabs-r@mail.nih.gov
Tissue Procurement: Dr. Steven Brown
CDNA Library Preparation: Dr. M. Bento Soares, University of Iowa
DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
Clone Distribution: Clone distribution information can be obtained
from Dr. M. Bento Soares, bento-soares@uiowa.edu
The following repetitive elements were found in this cDNA
sequence: 44-82, >AT-rich#low_complexity
Seq primer: M13 FORWARD
POLYA=yes.
Location/Qualifiers
1..520
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="UI-1-BC1p-ayx-c-01-0-UI"
/clone_lib="NCI CGAP_P13"
/tissue_type="Placenta"
/dev_stage="8-9 weeks"
/lab_host="DH10B (Life Technologies)"
/note="Organ: Placenta; Vector: pRTT3-Pac (Pharmacia) with
a modified polylinker, Site_1: EcoR I, Site_2: Not I;
NCI CGAP_P13 is a subcloned cDNA library constructed
according to Bonaldo, Lennon and Soares, Genome Research,
6:791-806, 1996. First strand cDNA synthesis was primed
with an oligo-dT primer containing a Not I site. Double
stranded cDNA was ligated to an EcoR I adaptor, digested
with Not I, and cloned directionally into pRTT3-Pac
vector. The oligonucleotide used to prime the synthesis of
first-strand cDNA contains a library tag sequence that is
located between the Not I site and the (dT)18 tail. The
sequence tags for this library are GA, AGGA. For
additional information, contact: Bento Soares,
bento-soares@uiowa.edu
TAG_LIB=UI-1-BC1p
TAG_TISSUE=Placenta human 8 week
TAG_SEO=GA"

BASE COUNT 127 a 147 c 115 g 131 t
 ORIGIN
 Query Match 37.2%; Score 520; DB 14; Length 520;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 520; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 867 ACTGTGTTCAATAGTGTGAGATGTAGTAAACCCCTGCTGCTGCTGCAATGCCA 926
 DB 520 ACTGTGTTCAATAGTGTGAGATGTAGTAAACCCCTGCTGCTGCTGCAATGCCA 461
 QY 927 CACAGAGCGGTGGGGGCTGGTGGGAGCATTCATGATGGAGTGTCTCTCAGCTTAGGT 986
 DB 460 CACAGAGCGGTGGGGGCTGGTGGGAGCATTCATGATGGAGTGTCTCTCAGCTTAGGT 401
 QY 987 CTGAGACAGAGACTTGGCGGGGAGTGTCTCCAGAGTGTGGTGAATTCGTACTGGGAGG 1046
 DB 400 CTGAGACAGAGACTTGGCGGGGAGTGTCTCCAGAGTGTGGTGAATTCGTACTGGGAGG 341
 QY 1047 CTATCTGTACCTCCGACAGGGGACACTCCAGGCCAGCCAGGGGTCAAGGGGAGAG 1106
 DB 340 CTATCTGTACCTCCGACAGGGGACACTCCAGGCCAGCCAGGGGTCAAGGGGAGAG 281
 QY 1107 TGACACCTCAGATGAGGCCAAGACTGGGGGTCAAGGGGAGAGTGTGGTGGAGCCAGAC 1166
 DB 280 TGACACCTCAGATGAGGCCAAGACTGGGGGTCAAGGGGAGAGTGTGGTGGAGCCAGAC 221
 QY 1167 CTGGGGCGGGGGGTGGGGGCGGGGCTTCTGCTCATTTGCTTCAATGAAAGCTCAAA 1226
 DB 220 CTGGGGCGGGGGGTGGGGGCGGGGCTTCTGCTCATTTGCTTCAATGAAAGCTCAAA 161
 QY 1227 GCAGCCAAAACAGAGCTTTTCCCTTCTGAGTTGAATATCCAGAACTTTTGTACTT 1286
 DB 160 GCAGCCAAAACAGAGCTTTTCCCTTCTGAGTTGAATATCCAGAACTTTTGTACTT 101
 QY 1287 CTGTGTGTTAAATGTTTATTTTGTAAATAATTAATTAATTAATTAATTAATTAAT 1346
 DB 100 CTGTGTGTTAAATGTTTATTTTGTAAATAATTAATTAATTAATTAATTAATTAAT 41
 QY 1347 GTTTCACAGCAAACTCTCCCTAAATAAAAAAAAAAAAAAAAAAAAAA 1386
 DB 40 GTTTCACAGCAAACTCTCCCTAAATAAAAAAAAAAAAAAAAAAAAAA 1

RESULT 4
 A1669816/c 529 bp mRNA linear EST 15-DEC-1999
 LOCUS tuj1f08.x1 NCI CGAP_P128 Homo sapiens cDNA clone IMAGE:2252679 3',
 DEFINITION mRNA sequence.
 ACCESSION A1669816
 VERSION A1669816.1 GI:4834590
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 REFERENCE 1 (bases 1 to 529)
 NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.
 TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
 Tumor Gene Index
 JOURNAL Unpublished (1997)
 COMMENT Contact: Robert Strausberg, Ph.D.
 Email: cgaps-remail.nih.gov
 Tissue Procurement: Michael J. Brownstein, M.D., Ph.D., Michael R.
 Emmert-Buck, M.D., Ph.D.
 cDNA Library Preparation: M. Bento Soares, Ph.D.
 cDNA Library Arrayed by: Greg Lennon, Ph.D.
 DNA Sequencing by: Washington University Genome Sequencing Center
 Clone distribution: NCI-CGAP clone distribution information can be
 found through the I.M.A.G.E. Consortium/UM/UMC at:
www-bio.lnli.gov/bbrp/image/image.html
 Insert length: 912 Std Error: 0.00
 Seq primer: -40UP from Gibco

FEATURES
 source
 Location/Qualifiers
 1..529
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="IMAGE:2252679"
 /clone_id="NCI_CGAP_P128"
 /sex="male"
 /dev_stage="adult"
 /lab_host="DH10B"
 /note="Organ: prostate; Vector: pTV73D-Pac (Pharmacia)
 with a modified polylinker; Plasmid DNA from the
 normalized library NCI CGAP_P128 was prepared, and ss
 circles were made in vitro. Following HAP purification,
 this DNA was used as tracer in a subtractive hybridization
 reaction. The driver was PCR-amplified cDNAs from a pool
 of 5,000 clones made from the same library (clonoids
 985608-986759, 1101192-1101959, and 1217928-1220615).
 Subtraction by Bento Soares and M. Fatima Bonaldo."

BASE COUNT 131 a 159 c 125 g 113 t
 ORIGIN
 Query Match 37.0%; Score 517; DB 9; Length 529;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 517; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 840 TGGAGCGTTCGATGGGAGAGCCCATCACTGTTCAATAGTGTAGATGAGCTAAAG 899
 DB 517 TGGAGCGTTCGATGGGAGAGCCCATCACTGTTCAATAGTGTAGATGAGCTAAAG 458
 QY 900 CCCCTGCTGCTGCTGCTGACATGACAGAGCGGTGGGGGCTGCGTGGGAGCAATCC 959
 DB 457 CCCCTGCTGCTGCTGCTGACATGACAGAGCGGTGGGGGCTGCGTGGGAGCAATCC 398
 QY 960 ATCGTAGAGTGTCTCTCAGCTTAGGTGAGTGTGACAGAGACTTGGCGGGGAGTGTCCAG 1019
 DB 397 ATCGTAGAGTGTCTCTCAGCTTAGGTGAGTGTGACAGAGACTTGGCGGGGAGTGTCCAG 338
 QY 1020 ATGTGGTGATTTCTGACCTGGGAGGGGCTATCTGACCTCCGACAGGGGACATCCCA 1079
 DB 337 ATGTGGTGATTTCTGACCTGGGAGGGGCTATCTGACCTCCGACAGGGGACATCCCA 278
 QY 1080 GGCAGCCAGGAGGTGACAGGGGAGAGGTGACACACCTCAGCATGAGCCAAAGCTGGGGTCA 1139
 DB 277 GGCAGCCAGGAGGTGACAGGGGAGAGGTGACACACCTCAGCATGAGCCAAAGCTGGGGTCA 218
 QY 1140 GGGAGCAGGTGTGTTTGTAGCCAGACCTGGGGGCGGGGCTTCTCTGACC 1199
 DB 217 GGGAGCAGGTGTGTTTGTAGCCAGACCTGGGGGCGGGGCTTCTCTGACC 158
 QY 1200 TCATTTGCTTTCAATGAAGAGCTCAAAAGCAGCCAAACAGAGCTTTCCCTTCTCGAG 1259
 DB 157 TCATTTGCTTTCAATGAAGAGCTCAAAAGCAGCCAAACAGAGCTTTCCCTTCTCGAG 98
 QY 1260 TTGAATATCCAGAACTTTTGTACTTCTTGTGTTAAATGTTTATTTTGTAAAAA 1319
 DB 97 TTGAATATCCAGAACTTTTGTACTTCTTGTGTTAAATGTTTATTTTGTAAAAA 38
 QY 1320 TAAATATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1356
 DB 37 TAAATATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1

RESULT 5
 BE250262 770 bp mRNA linear EST 13-JUL-2000
 LOCUS BE250262
 DEFINITION 600943273F1 NIH_MGC_17 Homo sapiens cDNA clone IMAGE:2959770 5',
 mRNA sequence.
 ACCESSION BE250262
 VERSION BE250262.1 GI:9120370
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
1 (bases 1 to 770)
NIH-MGC <http://imgc.nci.nih.gov/>
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgabbs-remail.nih.gov
Tissue Procurement: ATCC
CDNA Library Preparation: Ling Hong/Rubin Laboratory
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LINL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LINL at: imgc.lnl.gov
Plate: LHCMS2 row: a column: 19
High quality sequence stop: 730.

FEATURES

source

1..770
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:2959770"
/clone_lib="NIH MGC 17"
/tissue_type="rhodomyosarcoma"
/lab_host="DH10B (phage-resistant)"
/note="Organ: muscle; Vector: pOTB7; Site 1: EcoRI; Site 2: XhoI; CDNA made by oligo-dT priming. directionally cloned into EcoRI/XhoI sites using the following 5' adaptor: GGCACGAG(C). Size-selected >500bp for average insert size 1.8kb. Library constructed by Ling Hong in the laboratory of Gerald M. Rubin (University of California, Berkeley) using ZAP-CDNA synthesis kit (Stratagene) and Superscript II RT (Life Technologies)."

BASE COUNT

157 a 219 c 226 g 168 t

ORIGIN

Query Match 36.8%; Score 514; DB 10; Length 770;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 514; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

442 CAGGTGGGAGGCTGCTTTTCAAGATGATGAGGAGCCAGGTCCTGATCCCAACA 501
Db CAGGTGGGAGGCTGCTTTTCAAGATGATGAGGAGCCAGGTCCTGATCCCAACA 244
502 AGACCAATATGTAAGGCTCTGCTGAGCTTATCTGAGGGCTGGCTGACCACTGACT 561
Db AGACCAATATGTAAGGCTCTGCTGAGCTTATCTGAGGGCTGGCTGACCACTGACT 304
562 ATCTCAGAGAGCTGGGCTTCTGCTGAGGAGAGTGACTTGACCTGACAGCATGATGTC 621
Db ATCTCAGAGAGCTGGGCTTCTGCTGAGGAGAGTGACTTGACCTGACAGCATGATGTC 364
622 ACCGTGGGAGCCCTGAGAGCAAGCTAATCTCCAGACAGACAGATGTGACCGACGACAA 681
Db ACCGTGGGAGCCCTGAGAGCAAGCTAATCTCCAGACAGACAGATGTGACCGACGACAA 424
682 CGTGCAATATGCAATATGTAATGTAATGTAATGTAATGTAATGTAATGTAATGTAATG 741
Db CGTGCAATATGCAATATGTAATGTAATGTAATGTAATGTAATGTAATGTAATGTAATG 484
742 TCTAGTCCAGAGATGAGGAGGATGAGCTGCTTCCAACTCTGAGGGCTGTAAACAG 801
Db TCTAGTCCAGAGATGAGGAGGATGAGCTGCTTCCAACTCTGAGGGCTGTAAACAG 544
802 CTCAGAGCTAGTCTCCCACTGGGGGCTGTGCTCCCTCCCTGAGGAGGTTCCGTGGGACGC 861
Db CTCAGAGCTAGTCTCCCACTGGGGGCTGTGCTCCCTCCCTGAGGAGGTTCCGTGGGACGC 604
862 CCATCACTGTGTTCAATAGTGTGAGATGTGTAAGCCCTGCTGCTGCTGCTGCTGCTGCTG 921
Db CCATCACTGTGTTCAATAGTGTGAGATGTGTAAGCCCTGCTGCTGCTGCTGCTGCTGCTG 664
922 TGCACAGAGAGCGGTGGGGCTGCGTGGGAGCA 955

Db 665 TGCACAGAGAGCGGTGGGGCTGCGTGGGAGCA 698

RESULT 6
LOCUS BM671384/c 773 bp mRNA linear EST 27-FEB-2002
DEFINITION UI-E-CX1-afk-m-06-0-UI.s2 UI-E-CX1 Homo sapiens CDNA clone
UI-E-CX1-afk-m-06-0-UI.3', mRNA sequence.
BM671384
BM671384.1 GI:18981282
EST.

REFERENCE
AUTHORS
TITLE
JOURNAL
MEDLINE
COMMENT
Contact: Soares, MB
Program for Rat Gene Discovery and Mapping
University of Iowa
451 Eckstein Medical Research Building Iowa City, IA 52242, USA
Tel: 319 335 8250
Fax: 319 335 9565
Email: msoares@blue.weeg.uiowa.edu
Tissue Procurement: Dr. Gregg Hageman
CDNA library preparation: Dr. M. Bento Soares, University of Iowa
CDNA library Arrayed by: Dr. M. Bento Soares, University of Iowa
DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
Clone Distribution: Researchers may obtain clones from Research Genetics (www.resgen.com).
The following repetitive elements were found in this CDNA sequence: 1-59, >(TAAA)n#Simple_repeat (matched complement)
Seq primer: M13 Forward
POLYA=yes.

FEATURES

source

Location/Qualifiers

1..773
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="UI-E-CX1-afk-m-06-0-UI"
/clone_lib="UI-E-CX1"
/tissue_type="Retina Foveal and Macular"
/dev_stage="adult"
/lab_host="DH10B (Life Technologies) (T1 phage resistant)"
/note="Organ: eye; Vector: pT73-Pac (Pharmacia) with a modified polylinker; Site 1: EcoR I; Site 2: Not I; UI-E-CX1 is a normalized CDNA library containing the following tissue(s): Retina Foveal and Macular. The library was constructed according to Bonaldo, Lennon and Soares, Genome Research, 6:791-806, 1996. First strand CDNA synthesis was primed with an oligo-dT primer containing a Not I site. Double stranded CDNA was ligated to an EcoR I adaptor, digested with Not I, and cloned directionally into pT73-Pac vector. The oligonucleotide used to prime the synthesis that is located between the Not I site and the (dT)18 tail. The sequence tag for this library is GTCC. This library was created for the program, Gene Discovery in the Visual System, supported by National Eye Institute (NEI).
TAG_LIB=UI-E-CX1
TAG_TISSUE=Foveal and Macular Retina
TAG_SEQ=GTCC"

BASE COUNT

181 a 221 c 180 g 189 t 2 others

ORIGIN

Query Match 36.7%; Score 513; DB 14; Length 773;
Best Local Similarity 99.8%; Pred. No. 0;
Matches 563; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```

QY 780 AACCTGTGGGCTGTAGCAAGCTCAGGCTAGTCTCCCACTGGGGGCTGTGCCCCCTGCC 839
DB 595 AACCTGTGGGCTGTAGCAAGCTCAGGCTAGTCTCCCACTGGGGGCTGTGCCCCCTGCC 536
QY 840 TGGGAGCGTCCGCTGGGCAAGCCCATCATCTGTCAATAGTGAAGTGTAGTAAAG 899
DB 535 TGGGAGCGTCCGCTGGGCAAGCCCATCATCTGTCAATAGTGAAGTGTAGTAAAG 476
QY 900 CCCCTGTGCTGTGTGTGTCATGTCACAGCAGGCGGTGGGGCTGTGGGGCAATCC 959
DB 475 CCCCTGTGCTGTGTGTGTCATGTCACAGCAGGCGGTGGGGCTGTGGGGCAATCC 416
QY 960 ATGTGAGAGTGTCTCTCAGCTTGTGAGTGTGAGAGAGTGTGGCGGGGATGTCTCCAG 1019
DB 415 ATGTGAGAGTGTCTCTCAGCTTGTGAGTGTGAGAGAGTGTGGCGGGGATGTCTCCAG 356
QY 1020 ATGTGAGTGTATCTGTACCTGTGGGAGGCTATCTCTGACCTCCGACAGGGGACATCCCA 1079
DB 355 ATGTGAGTGTATCTGTACCTGTGGGAGGCTATCTCTGACCTCCGACAGGGGACATCCCA 296
QY 1080 GGGCAGCCAGGCGGTCTGAGGGGCAAGAGTGTGACACCTTGTGAGTGTGAGTGTGAGTGTG 1139
DB 295 GGGCAGCCAGGCGGTCTGAGGGGCAAGAGTGTGACACCTTGTGAGTGTGAGTGTGAGTGTG 236
QY 1140 GGGAGCAGTGTGTGTGTGAGCAGACCTGTGGGGCGGGGGCTGTGGGGCTGTGGGGCTGTGGGG 1199
DB 235 GGGAGCAGTGTGTGTGTGAGCAGACCTGTGGGGCGGGGGCTGTGGGGCTGTGGGGCTGTGGGG 176
QY 1200 TCATTGTCTTCAATGAAAGCCCTCAAGAGCAGCCAAACAGGCTTTTCCCTTCTCTGAG 1259
DB 175 TCATTGTCTTCAATGAAAGCCCTCAAGAGCAGCCAAACAGGCTTTTCCCTTCTCTGAG 116
QY 1260 TTGGAATATCCAGATCTTTTGTACTTCTGTGTGTGTAAATGTGTATTTTGTAAATAA 1319
DB 115 TTGGAATATCCAGATCTTTTGTACTTCTGTGTGTGTAAATGTGTATTTTGTAAATAA 56
QY 1320 TAAATATAAATAGTATAATAAT 1343
DB 55 TAAATATAAATAGTATAATAAT 32

RESULT 7
LOCUS BM668538/c 610 bp mRNA linear EST 27-FEB-2002
DEFINITION UI-E-CK1-afm-c-23-0-UI.s2 UI-E-CK1 Homo sapiens cDNA clone
ACCESSION BM668538
VERSION UI-E-CK1-afm-c-23-0-UI 3, mRNA sequence.
KEYWORDS BM668538.1 GI:18976369
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
REFERENCE 1 (bases 1 to 610)
AUTHORS Bonaldo,M.F., Lennon,G. and Soares,M.B.
TITLE Normalization and subtraction: two approaches to facilitate gene
discovery
JOURNAL Genome Res. 6 (9), 791-806 (1996)
MEDLINE 97044477
COMMENT Contact: Soares, MB
Program for Rat Gene Discovery and Mapping
University of Iowa
451 Eckstein Medical Research Building Iowa City, IA 52242, USA
Tel: 319 335 8250
Fax: 319 335 9565
Email: msoares@iuiiueeeg.uioa.edu
Tissue Procurement: Dr. Gregg Hageman
CDNA Library preparation: Dr. M. Bento Soares, University of Iowa
CDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa
DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
Clone Distribution: Researchers may obtain clones from Research
Genetics (www.resgen.com)
The following repetitive elements were found in this cDNA

```

```

FEATURES
SOURCE
Location/Qualifiers
1..610
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="UI-E-CK1-afm-c-23-0-UI"
/clone_lib="UI-E-CK1"
/tissue_type="Retina Foveal and Macular"
/dev_stage="adult"
/lab_host="DH10B (Life Technologies) (T1 phage resistant)"
/notes="Organ: Vector: pT73-Pac (Pharmacia) with a
modified polylinker; Site 1: EcoR I; Site 2: Not I;
UI-E-CK1 is a normalized cDNA library containing the
following tissue(s): Retina Foveal and Macular. The
library was constructed according to Bonaldo, Lennon and
Soares, Genome Research, 6:791-806, 1996. First strand
cDNA synthesis was primed with an oligo-dT primer
containing a Not I site. Double stranded cDNA was ligated
to an EcoR I adaptor, digested with Not I, and cloned
directionally into pT73-Pac vector. The oligonucleotide
used to prime the synthesis of first-strand cDNA contains
a library tag sequence that is located between the Not I
site and the (dr)18 tail. The sequence tag for this
library is GTCC. This library was created for the program,
Gene Discovery in the Visual System, supported by National
Eye Institute (NEI).
TAG LIB=UI-E-CK1
TAG TISSUE=FOveal and Macular Retina
TAG SEQ=GTCC"

BASE COUNT 145 a 147 g 143 t 1 others
ORIGIN
Query Match 35.3%; Score 493; DB 13; Length 610;
Best Local Similarity 99.7%; Pred. No. 0;
Matches 593; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 765 TATGACTGCTCTCCAAACCTGTGGGCTGTAGCAAGCTCAGGCTAGTCTCCCACTGGG 824
DB 610 TATGACTGCTCTCTCCAAACCTGTGGGCTGTAGCAAGCTCAGGCTAGTCTCCCACTGGG 551
QY 825 GGCTGTGCCCCCTCCCTGGGACGCTTCGTTGGGAGCCGCCCATCATCTGTGTCAATAGTGTG 884
DB 550 GGCTGTGCCCCCTCCCTGGGACGCTTCGTTGGGAGCCGCCCATCATCTGTGTCAATAGTGTG 491
QY 885 AGAATGTACTTAAAGCCCTCTGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 944
DB 490 AGAATGTACTTAAAGCCCTCTGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 431
QY 945 GGGTGGGACAAATCCATGTGTGAGTGTCTCTCAGCTTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1004
DB 430 GGGTGGGACAAATCCATGTGTGAGTGTCTCTCAGCTTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 371
QY 1005 GGGGAGTGTCTCAGAGATGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1064
DB 370 GGGGAGTGTCTCAGAGATGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 311
QY 1065 CAGGGGACACTCCCAAGCCCAAGGCTCAGGGGCAAGAGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1124
DB 310 CAGGGGACACTCCCAAGCCCAAGGCTCAGGGGCAAGAGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 251
QY 1125 CCAAGACTGGGGTCAAGGACAGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1184
DB 250 CCAAGACTGGGGTCAAGGACAGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 191
QY 1185 CCGGGGCTTTTGTGCTCTATTTGTTTCAATGAAGCTTCAAGACCCAAACACAGGCTT 1244
DB 190 CCGGGGCTTTTGTGCTCTATTTGTTTCAATGAAGCTTCAAGACCCAAACACAGGCTT 131
QY 1245 TCCCTCTCTGTGAGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1304
DB 130 TCCCTCTCTGTGAGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 71

```

QY 1305 TATTTTGTAAATAAATAAATAGTAAATAATGATGTTTCACAGCAA 1359
 DB 70 TATTTTGTAAATAAATAAATAAATAGTAAATAATGATGTTTCACAGCAA 16

RESULT 8
 BM142660/c 543 bp mRNA linear EST 29-NOV-2001
 LOCUS 1926C02.x1 Human Fetal Pancreas 1B Homo sapiens cDNA 3', mRNA
 DEFINITION sequence.
 ACCESSION BM142660.1 GI:117152727
 VERSION BM142660.1
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 REFERENCE 1 (bases 1 to 543)
 AUTHORS Melton, D., Brown, J., Kenty, G., Pernutt, A., Lee, C., Kaestner, R.,
 Lemishka, I., Searce, M., Brestelli, J., Gradwohl, G., Clifton, S.,
 Hillier, L., Marra, M., Pape, D., Wylie, T., Martin, J., Biststein, A.,
 Schmitt, A., Theising, B., Richter, E., Komko, I., Benne, J., Cardenas,
 M., Gibbons, M., McCann, R., Cole, R., Tsagaris, R., Williams, T.,
 Jackson, Y., and Bowers, Y.
 Endocrine Pancreas Consortium
 Unpublished (2000)
 CONTACT: Douglas Melton, Klaus H. Kaestner, & Hiroshi Inoue
 Endocrine Pancreas Consortium
 Harvard University, Howard Hughes Medical Institute
 Dept of Molecular and Cellular Biology, 7 Divinity Ave, Cambridge,
 MA 02138
 Tel: 617-495-1812
 Fax: 617-495-8557
 Email: dmelton@chp.harvard.edu
 Library was constructed by Dr. Hiroshi Inoue DNA sequencing by:
 Washington University Genome Sequencing Center For information on
 obtaining a clone please contact: Dr. Hiroshi Inoue
 (hinoue@imgate.wustl.edu)
 Seq primer: -40UP from Gibco
 High quality sequence stop: 445.
 Location/Qualifiers
 1. 543
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone_lib="Human Fetal Pancreas 1B"
 /issue_type="Fetal Pancreas (4 Pooled Donors, 18 - 20
 weeks, StrataGene #738023)"
 /dev_stage="Fetal Pancreas"
 /note="Vector: pBluescript SK(-); Site 1: NotI; Site 2:
 XhoI; cDNA made by oligo-dT priming. Size-selected on
 agarose gel. Average insert size ~1kb. 5' XhoI site was
 destroyed after directional cloning. Amplified once.
 Contact information: Hiroshi Inoue, MD, Metabolism Div.
 (Alan Permut Lab), Washington University School of
 Medicine, Box 8127, 660 South Euclid Ave., St. Louis, MO
 63110 E-mail: hinoue@imgate.wustl.edu Tel: 314-362-1916,
 Fax: 314-747-2692."

BASE COUNT 133 a 163 c 132 g 115 t

ORIGIN

Query Match 35.2%; Score 492; DB 13; Length 543;
 Best Local Similarity 99.8%; Pred. No. 0;
 Matches 542; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 814 TCCCACTGGGGCTGTGCTCCCTCTGAGACGTTCCGTGGACAGCCCATCATCTGTGT 873
 DB 543 TCCCACTGGGGCTGTGCTCCCTCTGAGACGTTCCGTGGACAGCCCATCATCTGTGT 484

QY 874 TCAATGTGTGAGAAATGAGCTAAAGCCCTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 933
 DB 483 TCAATGTGTGAGAAATGAGCTAAAGCCCTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 424

QY 934 CGGTGGGGCTGCGTGGGACAAATCCATGTCGAGCTTCTCGAGCTTAGCTTGACA 993
 DB 423 CGGTGGGGCTGCGTGGGACAAATCCATGTCGAGCTTCTCGAGCTTAGCTTGACA 364

QY 994 GGAGACTTGGCGGGGATGCTCCAGAGATGTGGTATTTCTGTAACCTGGGGAGCTATCTC 1053
 DB 363 GGAGACTTGGCGGGGATGCTCCAGAGATGTGGTATTTCTGTAACCTGGGGAGCTATCTC 304

QY 1054 TGACCTCCCGACAGGAGCACTCCAGGCGAGCCAGGGGCTAGGGGCGAGAGTGACAC 1113
 DB 303 TGACCTCCCGACAGGAGCACTCCAGGCGAGCCAGGGGCTAGGGGCGAGAGTGACAC 244

QY 1114 CTCACATGAGCAAGACTGGGTCAGGAGCAGGTGTGTTGAGCCAGACCTGGGGC 1173
 DB 243 CTCACATGAGCAAGACTGGGTCAGGAGCAGGTGTGTTGAGCCAGACCTGGGGC 184

QY 1174 GGGGGTGGGGCGGGGCTTTCTGCTCTATTTGCTTTCAATGAAAGCTCAAGACGCA 1233
 DB 183 GGGGGTGGGGCGGGGCTTTCTGCTCTATTTGCTTTCAATGAAAGCTCAAGACGCA 124

QY 1234 AAACGAGCTTCCCTCTCGAGTTGATATCCGAATCTTTGACTTCTGTG 1293
 DB 123 AAACGAGCTTCCCTCTCGAGTTGATATCCGAATCTTTGACTTCTGTG 64

QY 1294 GTTAATGTTATTTTGTAAATAAATAAATTAATTAATTAATTAATTAATTAATTAAT 1353
 DB 63 GTTAATGTTATTTTGTAAATAAATAAATAAATTAATTAATTAATTAATTAATTAATTAAT 4

QY 1354 AGC 1356
 DB 3 AGC 1

RESULT 9
 BE677292/c 504 bp mRNA linear EST 08-SEP-2000
 LOCUS 7d83b01.x1 lupski_dorsal root ganglion Homo sapiens cDNA clone
 DEFINITION IMAGE:3279529.3', similar to, contains MER11, c3 MSRI repetitive
 element; mRNA sequence.
 ACCESSION BE677292.1 GI:10037833
 VERSION BE677292.1
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 REFERENCE 1 (bases 1 to 504)
 AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
 National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
 Tumor Gene Index
 Unpublished (1997)
 CONTACT: Robert Strausberg, Ph.D.
 Email: cgapbs-r@mail.nih.gov
 Tissue Procurement: Dr. James Lupski
 cDNA Library Preparation: Lupski Laboratory
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNI)
 DNA Sequencing by: Washington University Genome Sequencing Center
 Clone distribution: NCI-CGAP clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNI at:
 info@image.llnl.gov
 Seq primer: -40UP from Gibco
 High quality sequence stop: 409.
 Location/Qualifiers
 1. 504
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone_lib="IMAGE:3279529"
 /clone_lib="lupski_dorsal_root_ganglion"
 /sex="male"
 /issue_type="dorsal root ganglia"
 /dev_stage="adult, 36 yr"
 /lab_host="DH10B"
 /note="Vector: pCMV-SPORT6 (Life Technologies); Site 1:

Notcl; Site 2: SalI; cDNA made by oligo-dT priming. Directionally cloned using the following adaptors:
 5'-TGACCCAGCGGCG-3' and
 5'-GACTAGTCTAGTACGAGGCGCCCTT(15)-3'. Size selected > 1 kb for average insert length 1.7 kb. This is a primary library, non-amplified. Library constructed by Life Technologies and donated by J. Lupski, M.D./Ph.D. (Baylor College of Medicine) and is available through Life Technologies."

BASE COUNT

124 a 147 c 115 g 118 t

ORIGIN

Query Match 34.8%; Score 486; DB 10; Length 504;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 486; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 858 AGCCCATCATCTGTTTCAATAGTGTGAGAAATGATGAAAGCCCTGCTGCTGCTG 917
 DB 504 AGCCCATCATCTGTTTCAATAGTGTGAGAAATGATGAAAGCCCTGCTGCTGCTG 445
 QY 918 CACATGCCACAGAGCGGCTGCGTGGGGAATCCATCGTGAAGTGTCTTC 977
 DB 444 CACATGCCACAGAGCGGCTGCGTGGGGAATCCATCGTGAAGTGTCTTC 385
 QY 978 AGCTTAGTGTGAACAGAGACTTTGGCGGGGATGCTCCAGATGTGGGTGATTCTG 1037
 DB 384 AGCTTAGTGTGAACAGAGACTTTGGCGGGGATGCTCCAGATGTGGGTGATTCTG 325
 QY 1038 CTGGGAGGCTATCTGACCTCCCGACAGGGGACATCCCGAGCCAGCCAGGGGTGAG 1097
 DB 324 CTGGGAGGCTATCTGACCTCCCGACAGGGGACATCCCGAGCCAGCCAGGGGTGAG 265
 QY 1098 GGGGAGAGTGCAACCTCAGATGAGCAAGATGCGGGTCAAGAGAGAGAGTGTGTTG 1157
 DB 264 GGGGAGAGTGCAACCTCAGATGAGCAAGATGCGGGTCAAGAGAGAGAGTGTGTTG 205
 QY 1158 AGCCAGAGACTTGGGGCGGGGCTGGGGGCTTCTGCTCATTTGCTTTCATGAA 1217
 DB 204 AGCCAGAGACTTGGGGCGGGGCTGGGGGCTTCTGCTCATTTGCTTTCATGAA 145
 QY 1218 AGCTCAAGAGAGCAAAACAGAGCTTCCCTCTCTGAGTTGAAATCCAGAAATCT 1277
 DB 144 AGCTCAAGAGAGCAAAACAGAGCTTCCCTCTCTGAGTTGAAATCCAGAAATCT 85
 QY 1278 TTGTACTCTTGTGTTGTTAAATGTTATTTTGTAAAAATTAATAATTAAGTTAA 1337
 DB 84 TTGTACTCTTGTGTTGTTAAATGTTATTTTGTAAAAATTAATAATTAAGTTAA 25
 QY 1338 TAAAT 1343
 DB 24 TAAAT 19

RESULT 10

EM726293 543 bp mRNA linear EST 01-MAR-2002

LOCUS UI-E-EJ0-aih-d-23-0-UI.r1 UI-E-EJ0 Homo sapiens cDNA clone

DEFINITION UI-E-EJ0-aih-d-23-0-UI 5', mRNA sequence.

ACCESSION BM726293

VERSION BM726293.1 GI:19047626

KEYWORDS EST.

SOURCE human.

ORGANISM

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE 1 (bases 1 to 543)

AUTHORS Bonaldo, M.F., Lennon, G. and Soares, M.B.

TITLE Normalization and subtraction: two approaches to facilitate gene

JOURNAL. Genome Res. 6 (9), 791-806 (1996)

MEDLINE 97044477

COMMENT Program for Rat Gene Discovery and Mapping

FEATURES

source

University of Iowa
 451 Eckstein Medical Research Building Iowa City, IA 52242, USA
 Tel: 319 335 8250
 Fax: 319 335 9565
 Email: mesores@iue.weeg.uiowa.edu
 Tissue Procurement: Dr. Gregg Hageman
 cDNA Library preparation: Dr. M. Bento Soares, University of Iowa
 DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
 Clone Distribution: Researchers may obtain clones from Research Genetics (www.resgen.com).
 Seq primer: M13 Reverse.

Location/Qualifiers

1. 543
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="UI-E-EJ0-aih-d-23-0-UI"
 /clone_1ib="UI-E-EJ0"
 /tissue_type="fetal eyes, lens, eye anterior segment, optic nerve, retina, Retina Foveal and Macular, RPE and Choroid"
 /dev_stage="fetal and adult"
 /lab_host="DH10B (Life Technologies) (T1 phage resistant)"
 /note="Organ: eye; Vector: pRT3-Pac (Pharmacia) with a modified polylinker; Site 1: EcoR I; Site 2: Not I; UI-E-EJ0 is a subtracted cDNA library constructed according to Bonaldo, Lennon and Soares, Genome Research, 6:791-806, 1996. First strand cDNA synthesis was primed with an oligo-dT primer containing a Not I site. Double stranded cDNA was ligated to an EcoR I adaptor, digested with Not I, and cloned directionally into pRT3-Pac vector. The oligonucleotide used to prime the synthesis of first-strand cDNA contains a library tag sequence that is located between the Not I site and the (dT)18 tail. The sequence tags for this library are: fetal eyes, AGAATCAGA; lens, CGATTAGCA; eye anterior segment, ATGCCGAT; optic nerve, CCATTAGTG; retina, CCGCG; Retina Foveal and Macular, GTCC; RPE and Choroid, ACCCTA. This library was created for the program, Gene Discovery in the Visual System, supported by National Eye Institute (NEI)."

BASE COUNT

120 a 155 c 150 g 118 t

ORIGIN

Query Match 34.8%; Score 486; DB 14; Length 543;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 486; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 442 CAGGTGGGAGCTGCTTTTCAAGACTGATGAGAGCCAAATGCTCTGATCCCAACA 501
 DB 58 CAGGTGGGAGCTGCTTTTCAAGACTGATGAGAGCCAAATGCTCTGATCCCAACA 117
 QY 502 AGACCAATATGTGAAGGCTGCTGCTGATCTGAGAGGCTGCGCTGACAGCTGACT 561
 DB 118 AGACCAATATGTGAAGGCTGCTGCTGATCTGAGAGGCTGCGCTGACAGCTGACT 177
 QY 562 ATCTCAGAGCTGGGCTTCTGCTGAGAGGAGTGAATTGCACTGGCAGCACTGATGTC 621
 DB 178 ATCTCAGAGCTGGGCTTCTGCTGAGAGGAGTGAATTGCACTGGCAGCACTGATGTC 237
 QY 622 ACCTGGAAACCTCTGCAACAAGCTTAATCTCCAGACAGACAGATGTGACCAAGCAAA 681
 DB 238 ACCTGGAAACCTCTGCAACAAGCTTAATCTCCAGACAGACAGATGTGACCAAGCAAA 297
 QY 682 CGTGCAATATGTGAAGGCTGCTGCTGATCTGAGAGGCTGCGCTGACAGCTGACT 741
 DB 298 CGTGCAATATGTGAAGGCTGCTGCTGATCTGAGAGGCTGCGCTGACAGCTGACT 357
 QY 742 TCCTAGTCCAGAAATATGAGGCTATGATGCTCTTCAACCTGTGGCTGTAGCAAG 801
 DB 358 TCCTAGTCCAGAAATATGAGGCTATGATGCTCTTCAACCTGTGGCTGTAGCAAG 417
 QY 802 CTGAGGCTATCTTCCCACTGGGGGCTGGCCCTCTGGAGAGGTTCCGTGGGAGGC 861

Db 418 CTGAGCTGATCTCCCTGAGGCTGTGCTCCCTGCTGCGACGGTTCGTGGGACGC 477
QY 862 CCATCACTGTGTTCAATAGTGTAGATGTAAGCCCTGCTGCTGCTGACCA 921
Db 478 CCATCACTGTGTTCAATAGTGTAGATGTAAGCCCTGCTGCTGCTGACCA 537
QY 922 TGGCAC 927
Db 538 TGGCAC 543

RESULT 11
A1453409 492 bp mRNA linear EST 13-APR-1999
A1453409/c t37h12.x1 NCI_CGAP_Pan1 Homo sapiens cDNA clone IMAGE:2143751 3',
DEFINITION mRNA sequence.
ACCESSION A1453409.1 GI:4281679
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 492)
NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
AUTHORS National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
TITLE Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Life Technologies catalog #: 11548-013
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
www.bio.llnl.gov/bbrp/image/image.html
Insert Length: 1208 Std Error: 0.00
Seq primer: -40UP from Glibco
High quality sequence stop: 438.
Location/Qualifiers
1. 492
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_1ib="NCI CGAP Pan1"
/tissue_type="adenocarcinoma"
/lab_host="DH10B"
/note="Organ: pancreas; Vector: pCMV-SPORT6; Site 1: SalI;
Site 2: NotI; Cloned unidirectionally. Primer: Oligo dt.
Average insert size 1.72-kb. Life Technologies catalog #:
11548-013"

BASE COUNT 123 a 146 c 112 g 111 t

Query Match 34.2%; Score 478; DB 9; Length 492;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 478; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 866 CACTGTGTTCAATAGTGTAGATGTAAGCCCTGCTGCTGCTGACACTGGC 925
Db 492 CACTGTGTTCAATAGTGTAGATGTAAGCCCTGCTGCTGCTGACACTGGC 433
QY 926 ACAGCAGGCGGTGGGGCTGCGTGGGACATCATCTGAGTGTCTCTCAGCTTAGG 985
Db 432 ACAGCAGGCGGTGGGGCTGCGTGGGACATCATCTGAGTGTCTCTCAGCTTAGG 373
QY 986 TCTGACAGAGACTTGGCGGGGATGCTTCAGAGATGTTGGTATTTCTGACTGGGAG 1045
Db 372 TCTGACAGAGACTTGGCGGGGATGCTTCAGAGATGTTGGTATTTCTGACTGGGAG 313
QY 1046 GCTATCTGACCTCCGACAGGGGACATCCAGGCCAGGCCAGGGGTCAAGGGAGAG 1105
Db 312 GCTATCTGACCTCCGACAGGGGACATCCAGGCCAGGCCAGGGGTCAAGGGAGAG 253

QY 1106 GTGACACCTCAGCATGAGACCAAGATGCGGGTCAAGGAGCAGTGTGTTGACCCAGA 1165
Db 252 GTGACACCTCAGCATGAGACCAAGATGCGGGTCAAGGAGCAGTGTGTTGACCCAGA 193
QY 1166 CTTGGGGGGGGGGTGGGGGCGGGGCTTTCTGCTCATTTGCTTTCAATGAAGCTCA 1225
Db 192 CTTGGGGGGGGGGTGGGGGCGGGGCTTTCTGCTCATTTGCTTTCAATGAAGCTCA 133
QY 1226 AGCAGCAAAAACAGGCTTTCCCTCTGAGTTGATATCCAGATCTTTTGTACT 1285
Db 132 AGCAGCAAAAACAGGCTTTCCCTCTGAGTTGATATCCAGATCTTTTGTACT 73
QY 1286 TCTTGTGTTAAATGTTTATTTTGAATAAATAAATAAATAGTTAAATAAT 1343
Db 72 TCTTGTGTTAAATGTTTATTTTGAATAAATAAATAAATAAATAGTTAAATAAT 15

RESULT 12
BQ012938 485 bp mRNA linear EST 26-MAR-2002
LOCUS BQ012938/c
DEFINITION UI-1-BC1P-ava-g-11-0-UI.s1 NCI CGAP Pl3 Homo sapiens cDNA clone
UI-1-BC1P-ava-g-11-0-UI 3', mRNA sequence.
ACCESSION BQ012938
KEYWORDS BQ012938.1 GI:19737839
SOURCE EST.
ORGANISM human.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 485)
NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
AUTHORS National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
TITLE Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: Dr. Steven Brown
cDNA Library preparation: Dr. M. Bento Soares, University of Iowa
cDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa
DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
Clone Distribution: Clone distribution information can be obtained
from Dr. M. Bento Soares, bento-soares@uiowa.edu
The following repetitive elements were found in this cDNA
sequence: 29-77, >AT rich#low_complexity
Seq primer: M13 FORWARD
PolyA=yes.
Location/Qualifiers
1. 485
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="UI-1-BC1P-ava-g-11-0-UI"
/clone_1ib="NCI CGAP Pl3"
/tissue_type="Placenta"
/dev_stage="8-9 weeks"
/lab_host="DH10B (Life Technologies)"
/note="Organ: Placenta; Vector: pRTT3-Pac (Pharmacia) with
a modified polylinker; Site 1: EcoR I; Site 2: Not II;
NCI CGAP Pl3 is a subcloned cDNA library constructed
according to Bonaldo, Lennon and Soares, Genome Research,
6:791-806, 1996. First strand cDNA synthesis was primed
with an oligo-dT primer containing a Not I site. Double
stranded cDNA was ligated to an EcoR I adaptor, digested
with Not I, and cloned directionally into pRTT3-Pac
vector. The oligonucleotide used to prime the synthesis of
first-strand cDNA contains a library tag sequence that is
located between the Not I site and the (dT)18 tail. The
sequence tags for this library are GA, AGGA. For
additional information, contact: Bento Soares,
bento-soares@uiowa.edu
TAG_LIB=UI-1-BC1P
TAG_TISSUE=Placenta human 8 week
TAG_SEQ=GA"

BASE COUNT 116 a 141 c 108 g 120 t

This clone is available royalty-free through LNL; contact the IMAGE Consortium (info@image.llnl.gov) for further information.
Seq primer: -40UP from Gibco
High quality sequence stop: 502.
Location/Qualifiers

FEATURES

source

1.522
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:2934556"
/clone.lib="Soares_NFL_T_GBC_S1"
/lab_host="DH10B"
/note="Organ: pooled; Vector: pRTT3D-Pac (Pharmacia) with a modified polylinker; Site 1: Not I; Site 2: Eco RI; Equal amounts of plasmid DNA from three normalized libraries (fetal lung NbHL19W, testis NHT, and B-cell NCI CGAP GC81) were mixed, and as circles were made in vitro. Following HAP purification, this DNA was used as tracer in a subtractive hybridization reaction. The driver was PCR-amplified cDNAs from pools of 5,000 clones made from the same 3 libraries. The pools consisted of I.M.A.G.E. clones 297480-302087, 682632-687239, 726408-728711, and 729096-731399. Subtraction by Bento Soares and M. Fatima Bonaldo." 1 others

BASE COUNT

129 a 156 c 121 g 115 t

ORIGIN

Query Match 33.4%; Score 467; DB 10; Length 522;

Best Local Similarity 99.8%; Pred. No. 0;

Matches 517; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

840 TGGGACGGTCCGTGGGACAGCCCATCATCTGTTCAATAGTGAATGACTAAG 899
518 TGGGACGGTCCGTGGGACAGCCCATCATCTGTTCAATAGTGAATGACTAAG 459
900 CCCCTGCTGCTGCTGCTGCACATGCCACAGACGGGCTGCTGGGACATCC 959
458 CCCCTGCTGCTGCTGCTGCACATGCCACAGACGGGCTGCTGGGACATCC 399
960 ATCGTAGAGTGTCTCTCAGACTTAGGCTTGAGACAGAGACTTGGCGGGGATGCTCCAG 1019
398 ATGTGAGAGTGTCTCTCAGACTTAGGCTTGAGACAGAGACTTGGCGGGGATGCTCCAG 339
1020 ATGTGAGAGTGTCTCTCAGACTTAGGCTTGAGACAGAGACTTGGCGGGGATGCTCCAG 1079
338 ATGTGAGAGTGTCTCTCAGACTTAGGCTTGAGACAGAGACTTGGCGGGGATGCTCCAG 279
1080 GGGCAGCCCAAGGGGTGAGGGGCGAGAGTGCACCTCAGCATAGCCAAAGCTGGGCTCA 1139
278 GGGCAGCCCAAGGGGTGAGGGGCGAGAGTGCACCTCAGCATAGCCAAAGCTGGGCTCA 219
1140 GGGAGAGAGTGTCTCTCAGACTTAGGCTTGAGACAGAGACTTGGCGGGGATGCTCCAG 1199
218 GGGAGAGAGTGTCTCTCAGACTTAGGCTTGAGACAGAGACTTGGCGGGGATGCTCCAG 159
1200 TCATTGCTTTCAATGAAAGCTCAAAAGCAGCCAAACAGAGCTTCCCTTCTCTGAG 1259
158 TCATTGCTTTCAATGAAAGCTCAAAAGCAGCCAAACAGAGCTTCCCTTCTCTGAG 99
1260 TTTGAATATTCAGAAATCTTTTGTACTCTTGTGTTGTTAAATGTTATTTTGTAAAAA 1319
98 TTTGAATATTCAGAAATCTTTTGTACTCTTGTGTTGTTAAATGTTATTTTGTAAAAA 39
1320 TAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 1357
38 TAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 1

RESULT 15

BM693225

LOCUS

BM693225

DEFINITION

UI-E-CK1-afm-c-23-0-UI.r1 UI-E-CK1 Homo sapiens cDNA clone

ACCESSION

BM693225

VERSION BM693225.1 GI:19006483
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 465)
Bonaldo,M.F., Lennon,G. and Soares,M.B.
Normalization and subtraction: two approaches to facilitate gene discovery

JOURNAL Genome Res. 6 (9), 791-806 (1996)
MEDLINE 97044477
COMMENT Contact: Soares, MB
Program for Rat Gene Discovery and Mapping
University of Iowa
451 Eckstein Medical Research Building Iowa City, IA 52242, USA
Tel: 319 335 8250
Fax: 319 335 9565
Email: msoares@blue.weeg.uiowa.edu

Tissue Procurement: Dr. Gregg Hageman
CDNA Library preparation by: Dr. M. Bento Soares, University of Iowa
CDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa
DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
Clone Distribution: Researchers may obtain clones from Research Genetics (www.resgen.com).
Seq primer: M13 Reverse.

FEATURES

source

Location/Qualifiers

1.465

/organism="Homo sapiens"

/db_xref="taxon:9606"

/clone="UI-E-CK1-afm-c-23-0-UI"

/clone.lib="UI-E-CK1"

/tissue.type="Retina Foveal and Macular"

/dev.stage="adult"

/lab_host="DH10B (Life Technologies) (T1 phage resistant)"

/note="Organ: eye; Vector: pRTT3-Pac (Pharmacia) with a modified polylinker; Site 1: EcoR I; Site 2: Not I; UI-E-CK1 is a normalized cDNA library containing the following tissue(s): Retina Foveal and Macular. The library was constructed according to Bonaldo, Lennon and Soares, Genome Research, 6:791-806, 1996. First strand cDNA synthesis was primed with an oligo-dT primer containing a Not I site. Double stranded cDNA was ligated to an EcoR I adaptor, digested with Not I, and cloned directionally into pRTT3-Pac vector. The oligonucleotide used to prime the synthesis of first-strand cDNA contains a library tag sequence that is located between the Not I site and the (dT)18 tail. The sequence tag for this library is GTCC. This library was created for the program, Gene Discovery in the Visual System, supported by National Eye Institute (NEI)."

Bye Institute (NEI)."

Bye Institute (NEI)."

Bye Institute (NEI)."

Bye Institute (NEI)."

Bye Institute (NEI)."

Bye Institute (NEI)."

Bye Institute (NEI)."

Bye Institute (NEI)."

Bye Institute (NEI)."

Bye Institute (NEI)."

Bye Institute (NEI)."

Bye Institute (NEI)."

Bye Institute (NEI)."

Bye Institute (NEI)."

Bye Institute (NEI)."

Bye Institute (NEI)."

Bye Institute (NEI)."

Bye Institute (NEI)."

Bye Institute (NEI)."

Bye Institute (NEI)."

Bye Institute (NEI)."

Bye Institute (NEI)."

Bye Institute (NEI)."

Bye Institute (NEI)."

Bye Institute (NEI)."

Bye Institute (NEI)."

Bye Institute (NEI)."

Bye Institute (NEI)."

Bye Institute (NEI)."

Bye Institute (NEI)."

Bye Institute (NEI)."

Bye Institute (NEI)."

Bye Institute (NEI)."

Bye Institute (NEI)."

Bye Institute (NEI)."

Bye Institute (NEI)."

Bye Institute (NEI)."

Bye Institute (NEI)."

Bye Institute (NEI)."

QY	753	GAATCATGGGGGATAGTACTGCTCTCTCCAAACCTGTGGGCTGTAAAGCAAGCTCAGGCTAGT	812
Db	241	GAATCATGGGGGATAGTACTGCTCTCTCCAAACCTGTGGGCTGTAAAGCAAGCTCAGGCTAGT	300
QY	813	CTCCCACTGGGGGCTGTGCCCCCTCCCTGGGAGGTTCCGTGGGAGCCCCCATCCTGTG	872
Db	301	CTCCCACTGGGGGCTGTGCCCCCTCCCTGGGAGGTTCCGTGGGAGCCCCCATCCTGTG	360
QY	873	TTCAATAGTGTGAGATGTAGTAAAGCCCTGCTGCTGCTGCTGCACATGCCACAGCAG	932
Db	361	TTCAATAGTGTGAGATGTAGTAAAGCCCTGCTGCTGCTGCTGCACATGCCACAGCAG	420
QY	933	GCAGTGGGGGCTGCGTGGGGACAATCCATGCTGAGTGTCTCTC	977
Db	421	GCAGTGGGGGCTGCGTGGGGACAATCCATGCTGAGTGTCTCTC	465

Search completed: June 29, 2003, 06:37:13
 Job time : 1522 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on: June 24, 2003, 13:41:02 ; Search time 82.5 seconds
(without alignments)
4512.754 Million cell updates/sec

Title: US-09-989-919-15

Perfect score: 2527
Sequence: 1 ggtgctgcacctctaccgga.....aaaaaaaaaagcgctc 1397

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 1816940

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame+ n2p.model -DEV=xlp
-O/cgcn2.1/USPTO.spool/US09989919/runat.24062003.102045.19166/app.query.fasta_1.1543
-DB=A.GeneSeq.101002 -QPMT=faetan -SUFFIX=rag -MINMATCH=0.1 -LOOPTCL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi
-LIST=45 -DOCALLIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=200000000
-USER=US09989919 @CGCN 1.114 @runat.24062003.102045.19166 -NCPU=6 -ICPU=3
-NO MAP -LARGEOUTERY -NEG SCORES=0 -WAIT -DSRBLCK=100 -LONLOC
-DEV TIMEOUT=120 -MARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

- 1: /SID2/gcgdata/geneSeq/geneSeq-emb1/AA1980.DAT:*
- 2: /SID2/gcgdata/geneSeq/geneSeq-emb1/AA1981.DAT:*
- 3: /SID2/gcgdata/geneSeq/geneSeq-emb1/AA1982.DAT:*
- 4: /SID2/gcgdata/geneSeq/geneSeq-emb1/AA1983.DAT:*
- 5: /SID2/gcgdata/geneSeq/geneSeq-emb1/AA1984.DAT:*
- 6: /SID2/gcgdata/geneSeq/geneSeq-emb1/AA1985.DAT:*
- 7: /SID2/gcgdata/geneSeq/geneSeq-emb1/AA1986.DAT:*
- 8: /SID2/gcgdata/geneSeq/geneSeq-emb1/AA1987.DAT:*
- 9: /SID2/gcgdata/geneSeq/geneSeq-emb1/AA1988.DAT:*
- 10: /SID2/gcgdata/geneSeq/geneSeq-emb1/AA1989.DAT:*
- 11: /SID2/gcgdata/geneSeq/geneSeq-emb1/AA1990.DAT:*
- 12: /SID2/gcgdata/geneSeq/geneSeq-emb1/AA1991.DAT:*
- 13: /SID2/gcgdata/geneSeq/geneSeq-emb1/AA1992.DAT:*
- 14: /SID2/gcgdata/geneSeq/geneSeq-emb1/AA1993.DAT:*
- 15: /SID2/gcgdata/geneSeq/geneSeq-emb1/AA1994.DAT:*
- 16: /SID2/gcgdata/geneSeq/geneSeq-emb1/AA1995.DAT:*
- 17: /SID2/gcgdata/geneSeq/geneSeq-emb1/AA1996.DAT:*
- 18: /SID2/gcgdata/geneSeq/geneSeq-emb1/AA1997.DAT:*
- 19: /SID2/gcgdata/geneSeq/geneSeq-emb1/AA1998.DAT:*
- 20: /SID2/gcgdata/geneSeq/geneSeq-emb1/AA1999.DAT:*
- 21: /SID2/gcgdata/geneSeq/geneSeq-emb1/AA2000.DAT:*
- 22: /SID2/gcgdata/geneSeq/geneSeq-emb1/AA2001.DAT:*
- 23: /SID2/gcgdata/geneSeq/geneSeq-emb1/AA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	519.5	20.6	540	22	Novel protein kina
2	431	17.1	365	22	Novel protein kina
3	314	12.4	482	22	Novel protein kina
4	168	6.6	63	22	Novel human diagno
5	156.5	6.2	1093	14	Human EST encoded
6	144.5	5.6	1040	22	Human myoconic dys
7	139.5	5.4	1335	22	Novel human diagno
8	134.5	5.3	738	23	Human BCL9 homolog
9	125	4.9	384	22	Novel human secret
10	124	4.8	1239	15	Glutamic acid rece
11	123.5	4.8	505	22	Novel human diagno
12	123.5	4.8	1212	20	Human N-methyl-D-a
13	122.5	4.8	647	22	Brassica napus PER
14	122.5	4.8	900	22	Novel human diagno
15	122	4.8	343	21	Corn Adenylsulph
16	121.5	4.7	522	22	Novel signal trans
17	121.5	4.7	524	22	Novel signal trans
18	121.5	4.7	1400	21	Human OREF ORP402
19	120.5	4.7	871	22	Human protein SEQ
20	120.5	4.7	871	22	Human protein SEQ
21	120	4.7	684	22	Novel human diagno
22	120	4.7	815	22	Drosophila melanog
23	120	4.7	1061	20	Human N-methyl-D-a
24	120	4.7	1664	20	Mus musculus notch
25	119.5	4.7	626	22	Human polyprotein
26	119.5	4.7	806	22	Novel human diagno
27	119.5	4.7	806	22	Novel human diagno
28	119.5	4.7	855	22	Angiotensin conver
29	119.5	4.7	1214	15	Human NMDAR2 recep
30	119.5	4.7	1219	15	Human NMDAR2 recep
31	119.5	4.7	1231	15	Human NMDAR2 recep
32	119.5	4.7	1336	15	Human N-methyl-D-a
33	119.5	4.7	1339	15	Human NMDAR2 recep
34	119.5	4.7	1244	15	Human NMDAR2 recep
35	119	4.6	749	22	Drosophila melanog
36	119	4.6	897	22	Cytokine receptor
37	119	4.6	897	22	Cytokine receptor
38	119	4.6	1063	16	Infectious rubella
39	118.5	4.6	542	22	Human secreted pro
40	118.5	4.6	1619	23	Human lipid metabo
41	118	4.6	897	13	Sequence of beta-c
42	118	4.6	897	22	Cytokine receptor
43	118	4.6	897	22	Cytokine receptor
44	118	4.7	961	23	Human TSP1 domain
45	118	4.6	1061	20	Human N-methyl-D-a

ALIGNMENTS

RESULT 1
AA65697 standard; Protein: 540 AA.
AC AA65697;
XX
DT 27-MAR-2001 (first entry)
XX
DE Novel protein kinase, SEQ ID NO: 225.
XX
KW Human; mouse; protein kinase; antiarthritic; antisclerotic; osteoparitic;
KW immunosuppressive; cardiac; renal; antiinflammatory; antiaschemic;
KW dermatological; antidiabetic; antifertility; gene therapy; vaccina;
KW immune disorder; cardiovascular disease; neurodegenerative disease;
KW cancer; autoimmune disorder; stroke; inflammatory bowel disease;
KW inflammatory pelvic disease; multiple sclerosis; psoriasis.
XX
OS Homo sapiens.
XX

PN MO200073469-A2.
 PD 07-DEC-2000.
 PF 26-MAY-2000; 2000MO-US14842.
 PR 28-MAY-1999; 99US-0136503.
 PA (SUGEN-) SUGEN INC.
 PI Plowman GD, Martinez R, Whyte D, Sudersanam S;
 DR WPI: 2001-032161/04.
 DR N-PSDB; AAF44725.
 XX Nucleic acids encoding kinase polypeptides, useful for diagnosing and
 PT treating immune-related diseases and disorders, cardiovascular disease,
 PT neurodegenerative diseases and/or cancers -
 PS Claim 10; Fig 1; 310pp; English.

XX The present sequence is a novel protein kinase. The novel protein kinases
 CC and the nucleic acids that encode them may be used in the treatment and
 CC diagnosis of diseases associated with inappropriate kinase expression
 CC such as immune-related diseases and/or cancers. The nucleic acids and
 CC neurodegenerative diseases and/or cancers. The nucleic acids and
 CC complementary sequences may also be used as DNA probes in diagnostic
 CC assays. The kinase polypeptides may be used as antigens in the production
 CC of antibodies of kinase expression and activity. Anti-kinase antibodies
 CC and kinase antagonists may also be used to down regulate kinase
 CC expression and activity. Diseases related to kinase expression and
 CC activity include rheumatoid arthritis, atherosclerosis, autoimmune
 CC disorders, complications of organ transplantation, myocardial infarction,
 CC immune disorders, cardiomyopathies, strokes, renal failure,
 CC oxidative-stress related disorders, chronic inflammatory bowel disease,
 CC chronic inflammatory pelvic disease, multiple sclerosis, asthma,
 CC osteoarthritis, psoriasis, rhinitis, autoimmunity, diabetes, cancers and
 CC reproductive disorders.

XX Sequence 540 AA;

XX Alignment Scores:

Pred. No.: 66-40 Length: 540
 Score: 519.50 Matches: 105
 Percent Similarity: 59.66% Conservative: 0
 Best Local Similarity: 59.66% Mismatches: 1
 Query Match: 20.56% Indels: 70
 DB: 22 Gaps: 1

US-09-989-919-15 (1-1397) x IAB65697 (1-540)

QY 2 GTGTGACCTGTATCCGAGGCGGACAGTATCTGACAGACTCCAGCGAGCAGACAGTACC 61
 DB 435 ValLeuHisLeuTyrArgSerGlyGlnTyrLeuIndInAsnSerThrAlaSerSerThr 454
 QY 62 GAGTACCAAGTATATCCAGACAGACATCCCGGAGAAACTACCGCTGCGCCATCC 121
 DB 455 GlnTyrGlnCysIleProAspSerThrIleProIndInAspTyrArgCysTyrProSer 474
 QY 122 TACACACAGCGAGGAGCTCCCTTCCTTCAAGTGTTCACCTGGGTGAGGTGTCTGT 181
 DB 475 TyrHisHisGlySerCysLeuLeuSerValPheHisLeuAlaGlnValAlaValCys 494
 QY 182 GAGAGCATGCTCCAGTGTGGGCTTTGTGTGTCACCAACAGACACCTGGACAGGTGAG 241
 DB 495 GluSerHisAlaGlnCysArgAlaPheValValThrAsnGlnThrThrTyrGlnArg 514
 QY 242 CCAATGGAGAGAGCCCTTCCAGAGGAATGACAGAGACTCTCTGAGAGTTGATGAGACT 301
 DB 515 Gln----- 515
 QY 302 GATCCCATCGAAGTACAGAGGGGTGCTGAGTGATGAGAGAGATATACGTGTCTT 361

DB 515 ----- 515
 QY 362 CAAGCAGTCAATATAGGAGAGATGTCTTGTGCTCCAGAAAGAGAAATCAGCCCTGT 421
 DB 515 ----- 515
 QY 422 TACTCTCATCCTTGCCCCCAGTGGCAGCTGCTTTTTCAGACTGTGATGAGCCA 481
 DB 516 -----
 QY 482 ACTGTCTCCCTGATCCCAACAGACCAATATGTGAAGGCTCTGCG 527
 DB 525 nValValProAspProAsnLysThrThrTyrValLysAlaSerGly 540
 RESULT 2
 AAB59576
 ID AAB59576 standard; Protein; 365 AA.
 AC AAB59576;
 XX 27-MAR-2001 (first entry)
 DT Novel protein kinase, SEQ ID NO: 226.
 DE Human; mouse; protein kinase; antiarthritic; antisclerotic; osteopathic;
 KW immunosuppressive; cardiac; renal; antiinflammatory; antiasthmatic;
 KW dermatological; antidiabetic; antifertility; gene therapy; vaccine;
 KW immune disorder; cardiovascular disease; neurodegenerative disease;
 KW cancer; autoimmune disorder; stroke; inflammatory bowel disease;
 KW inflammatory pelvic disease; multiple sclerosis; psoriasis.
 XX Mus musculus.
 OS
 XX WO200073469-A2.
 PN 07-DEC-2000.
 PD 26-MAY-2000; 2000MO-US14842.
 PF 28-MAY-1999; 99US-0136503.
 PR (SUGEN-) SUGEN INC.
 PA Plowman GD, Martinez R, Whyte D, Sudersanam S;
 DR WPI: 2001-032161/04.
 DR N-PSDB; AAF44726.
 XX Nucleic acids encoding kinase polypeptides, useful for diagnosing and
 PT treating immune-related diseases and disorders, cardiovascular disease,
 PT neurodegenerative diseases and/or cancers -
 PS Claim 10; Fig 1; 310pp; English.

XX The present sequence is a novel protein kinase. The novel protein kinases
 CC and the nucleic acids that encode them may be used in the treatment and
 CC diagnosis of diseases associated with inappropriate kinase expression
 CC such as immune-related diseases and disorders, cardiovascular disease,
 CC neurodegenerative diseases and/or cancers. The nucleic acids and
 CC complementary sequences may also be used as DNA probes in diagnostic
 CC assays. The kinase polypeptides may be used as antigens in the production
 CC of antibodies of kinase expression and activity. Anti-kinase antibodies
 CC and kinase antagonists may also be used to down regulate kinase
 CC expression and activity. Diseases related to kinase expression and
 CC activity include rheumatoid arthritis, atherosclerosis, autoimmune
 CC disorders, complications of organ transplantation, myocardial infarction,
 CC immune disorders, cardiomyopathies, strokes, renal failure,
 CC oxidative-stress related disorders, chronic inflammatory bowel disease,
 CC chronic inflammatory pelvic disease, multiple sclerosis, asthma,
 CC osteoarthritis, psoriasis, rhinitis, autoimmunity, diabetes, cancers and
 CC reproductive disorders.

XX Sequence 365 AA;

Alignment Scores:

Pred. No.:	1.09e-11.	length:	366
Score:	431.00	Matches:	90
Percent Similarity:	53.71%	Conservative:	4
Best Local Similarity:	51.43%	Mismatches:	10
Query Match:	17.06%	Indels:	71
DB:	22	Gaps:	2

US-09-989-919-15 (1-1397) X AAB59576 (1-365)

Oy	5	CTCCACCTGTAACGGAGCGGAGCATATGCAGAAGTCCAGGCAAGCAGCATGCCAG	64
Db	262	LeuHiIleuNpheNkrsrctLygInTyLreugImnserrTh-----SetserArgAlaGlu	280
Oy	65	TACCAGTGTAATCCCACAACGACACATCCCCCAGGAAGACTCCGTGTGGCCATCTTAC	124
Db	281	TyglHArgrgleProAspserrAlalThrngInuAperTyArgCystrProSerTyr	300
Oy	125	CACCACGGAGCGTCCTCCCTTCAGTGTTCACTGGCTGAGCGTGAGATGTCTGGAG	184
Db	301	HishIeglyglYysrIeuNeuSerValAPheAnsLeuAlaGluAlaIleAspValCyGlu	320
Oy	185	AGCCATGCCCAAGTGTGGGCGCTTTGTGGTCACCAAACGACCACTGGACAGGTGACCA	244
Db	321	SerHisIeAgInCysArgrAlaPheValAlaTrnAsengInThrTrpThgIyArg-Ly	340
Oy	245	GTTGGGAGAAGCCTTCCAAGGAGATGGCAGGACCTCTGGAGGTTGATAGATAGAT	304
Db	340	B-----	340
Oy	305	CCCCCATCGGAAGTCAGAGGGGGGTGCTGAGGTAGAGAGAGATATACGTGCTTTCAA	364
Db	340	-----	340
Oy	365	GGCAGTCAATTAGGAGAAATGSGTCTTGCCCTCCAGAAAAGAAAAATCACGCCCTGTAC	424
Db	340	-----	340
Oy	425	CTCTCACCTTGCCCGCCCAAGTGGCAGCTGAGTCTTTTTCAAGACTGATGAGCCAAAT	484
Db	341	-----LeuValAPhePheYsrTrgIyTrpAsnGlnVa	351
Oy	485	GGTCCCGTATCCCAAGAACCAACATATGTGAAGGCGCTCTGGC	527
Db	351	IValProAspAlaGlyLystrThrTrnTyValAllysAlaProGly	365

RESULT 3
ABCG14651

ID ABG14651 standard; Protein; 482 AA

AC ABG14651;

DT 18-FEB-2002 (first entry)

DE XX Novel] human diacylglycerol protease #14643

11

KW food supplement; medical imaging; diagnostic; genetic disorder

Homo sapiens

XX WO2001 75067-A2
PN

XX
:
9
2
3

XX

FF 30-MAR-2001; 2001MO-050603L
XX

PR 31-MAR-2000; 2000US-0540217.
DP 23-APR-2000; 2000US-0649167.

XXXXXX

XX

PI Drmanac RT, Liu C, Tang YH,
XX
DR WPI; 2001-639362/73.
DR N-PSDB; AAS78838.
XX

PT New isolated polynucleotide and encoded polypeptides, useful in diagnostics, forensics, gene mapping, identification of mutations responsible for genetic disorders or other traits and to assess biodiversity -

PS Claim 20; SEQ ID No 45010; 103pp; English;
xx

The invention relates to isolated polynucleotide (I) and polypeptide (II) sequences. (I) is useful as hybridisation probes, polymerase chain reaction (PCR) primers, oligomers, and for chromosome and gene mapping, and in recombinant production of (II). The polynucleotides are also used in diagnostics as expressed sequence tags for identifying expressed genes. (I) is useful in gene therapy techniques to restore normal activity of (II) or to treat disease states involving (II). (II) is useful for generating antibodies against it, detecting or quantifying a polypeptide in tissue, as molecular weight markers and as a food supplement. (II) and its binding partners are useful in medical imaging of sites expressing (II). (I) and (II) are useful for treating disorders involving aberrant protein expression or biological activity. The polypeptide and polynucleotide sequences have applications in diagnostics, forensics, gene mapping, identification of mutations responsible for genetic disorders or other traits to assess biodiversity and to produce other types of data and products dependent on DNA and amino acid sequences. ABG00010-ABG30377 represent novel human diagnostic amino acid sequences of the invention.

Note: The sequence data for this patent did not appear in the printed specification, but was obtained in electronic format directly from WIPO at [ftp.wipo.int/pub/published_pat_sequences](http://wipo.int/pub/published_pat_sequences).

Sequence	482 AA
----------	--------

Alignment Scores:	
Pred. NO.:	1.22e-20
Score:	314.00
Percent Similarity:	41.06%
Best Local Similarity:	34.31%
Query Match:	12.43%
DB:	22
	12
	13
	13
	48
Length:	48
Matches:	11
Conservative:	23
Mismatches:	25
Indels:	13
Gaps:	12

US-09-989-919-15. (1-1397) x ABG14651 (1-482)

Oy		TGCTGCACCTGTAAACGGAGCGGGGAG--TATCTGCAGAACCCTCAGGGCAAGACGA---GT	58
Db	224	CysCysThrCysThrIValIAlaIlyrIglYLeuGlnAanSerThIAAser-ThProVa	24
Oy	59	ACCGCATACCAAGTGTAATCCAGACAGACACATCCCCAG-GAAGACTACCCGCTGTGGC	11
Db	243	IProserThrSerValSerGlnThIAIProserProIySclumapryrArgCyStrPr	26
Oy	118	ATCTTACCAACCAAGGAG-CTGCTCTTTTCAGTGTCAACTGGCTGAGGCTGTGATG	176
Db	263	oSertyrHISglYserLeuProPheSerValInProgly**GIYcysglyme	283
Oy	177	-TCTGTGAAGACCATGCCAGTGTGGGCTTTGTGTGTCACCAACCAACAACCTGGACA	23
Db	283	tSerValAIaSGIAmetProSerValGIYProleuTrpserProThraGProproGIyl	303
Oy	236	GCGAGCACAGTGGAGAAAGCCTTCCAAGGAGAGATGGAGAACCTCTGTGAGGTGATA	295
Db	303	nile-----	30-
Oy	296	GATAGTAGATCCCATTCCGAGACTCAGAGGAGGTGCTGAGTGATGAGAGAGGTATACG	355
Db	305	-----LysArgGlnAanTyrlieu-----	310
Oy	356	TGCTTTCAAAGCAGTCAAATTGGGAGGATGTCTTGGCTCTCAGAAAGAAACATCAG	415
Db	311	-----Leu-AanTyrlTrpSerCys-----	316

	RESULT 4
ID	AAM23513
XX	AAM23513 standard; Protein; 63 AA.
AC	AAM23513;
DT	12-OCT-2001 (first entry)
DE	Human EST encoded protein SEQ ID NO: 1038.
KW	Human; sheep; pig; cow; fruit fly; Yeast; hamster; macaque; horse; tomato; monkey; dog; sea urchin; expressed sequence tag; EST; diagnostics; forensic test; gene mapping; genetic disorder; biodiversity; gene therapy; nutrition.
OS	Homo sapiens.
PN	WO200154477-A2.
PD	02-AUG-2001.
PF	25-JAN-2001; 2001MCO-US02687.
PR	25-JAN-2000; 2000US-0491404. 17-JUL-2000; 2000US-0617746. 03-AUG-2000; 2000US-0631451. 15-SEP-2000; 2000US-0663870.
PA	(HYSE-) HYSEQ INC.
PT	Tang YT, Liu C, Zhou P, Qian XB, Wang Z, Chen R, Asundi V, Cao Y, Dmanac RA, Zhang J, Werhman T;

XX WPI; 2001-476164/51.
DR N-PSDB; AAH98172.
XX
FT Isolated polypeptide for treatment of diseases, diagnostics, raising
PT antibodies and research use
XX
PS Claim 20; Page 795; 1275pp; English.
XX
CC The present invention provides the protein and coding sequences of novel
CC proteins from a variety of organisms, including human, dog, cat, horse,
CC cow, pig, hamster, monkey, macaque, yeast, bacteria, fruit fly, sea
CC urchin and tomato. These were derived from expressed sequence tags (ESTs
CC from the organism of interest. They can be used in diagnostics,
CC forensics, gene mapping, identification of mutations, to assess
CC biodiversity and for nutritional purposes. The present sequence is a
CC protein of the invention.
XX
SQ Sequence 63 AA;
XX
Alignment Scores:
Pred. No.: 3.09e-07 Length: 63
Score: 168.00 Matches: 30
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 6.56% Indels: 0
DB: 22 Gaps: 0
US-09-989-919-15 (1-1397) x AAM23513 (1-63)
QY 90 ATGGTGTCTGTGGGATACACTGGTACTGGTACTGCTTGGCCGGAGTTCTGCAGA 31
Db 1 MelValleuseretlyllehtstptlyrserValleuDeuAlaValGlnurhecysarg 20
QY 30 TACTGCCCGCTCCGGTACAGGTGCACGACC 1
Db 21 TyrcysrrolenargtyrArgcysserThr 30
RESULT 5
AAR41001
ID AAR41001 standard; Protein; 1093 AA.
XX
AC AAR41001;
XX
DT 25-FEB-1994 (first entry)
XX
DE Human myotonic dystrophy gene protein.
XX
DE Abnormality: muscular dystrophy; CHR 19; chromosome 19;
KW protein kinase; polymerase chain reaction; brain.
XX
OS Homo sapiens.
XX
OS
XX
Key
FH Region 1..1093
FT /note="encoded by predicted reading frame a,
FT x's in the sequence indicate stop codons
FT in the reading frame"
XX
PN W09317104-A.
XX
PD 02-SEP-1993.
XX
PF 19-FEB-1993; 93WO-US01545.
XX
PR 20-FEB-1992; 92US-0839255.
XX
PA (MAST) MASSACHUSETTS INST TECHNOLOGY.
XX
PI Brook JD, Housman DE;
XX
DR WPI; 1993-288410/36.
XX

PT DNA sequence of myotonic dystrophy gene - used to produce probes
 PT and identify CHR 19 abnormality and protein kinase responsible
 XX
 PS Disclosure; Fig 6; 64pp; English.

CC The sequence is that encoded by predicted reading frame a of
 CC the human myotonic dystrophy (DM) gene. It may be used in the
 CC identification of individuals affected by DM.

XX Sequence 1093 AA;

Alignment Scores:

Pred. No.:	Length:	Score:	Matches:	Conservative:	Best Local Similarity:	Mismatches:	Indels:	Gaps:
1.07e-05	1093	156.50	127	52	23.39%	162	202	28

US-09-989-919-15 (1-1397) x AAR41001 (1-1093)

```

QY 24 GGCAGATATGCG-----AGACTCCAGCGGACAGACAGTACCGAGTACAGTGA 74
Db 605 GlySerThrCysLeuLeuTrpAlaThrPro--ThrProAlaTrpProSerGlyThrVal 623
QY 75 TCCAGACAGACACATCCCGCC-----AGGAGACTACCGCTGCTGCGCATCTTACC 125
Db 624 ArgSerGlnAlaProHisProTrpLysTrpArgProSerSerCysLeuSerHisThrCys 643
QY 126 ACCAGCGAGCTGCTCTTTCAGTTCACCTGAGCTGAG-----CTGGAGATGCT 179
Db 644 LysArgProAlaTrpSerProArgCysProHisArgMetLysGlnLeuLysTrpGlnPhe 663
QY 180 GTCAGAGCCATGCCAGCTGTCGGCCTTTGTGTGTCACCAACAGACACCTGACAGG 239
Db 664 GlnArgLeu-----SerLeuArgGlnArgLeuArgProArg***ArgCys 678
QY 240 AGCAGTGGGAGAGCCCTTCCA-----GGGAGA 269
Db 679 GlySerSerArgLysProTrpArgArgCysSerProGlyArgAla***AlaGlyArg 698
QY 270 TGGCAGACCTCTCTGAGGTGATAGATGATGATCCCATC-----GGAAGTCAG 323
Db 699 Trp-ArgProSerAlaArgThrThrArg--ThrSerProValAsnTyrAlaArgGln 717
QY 324 GGGGTGCTGAGTGAAGAGAGAGATACGTCTTCAAGCAGTCAATTAAGGAGA 383
Db 717 GlnGlyThrGlyThr***ArgHisThrSerGlySerCysArgSerGly----- 733
QY 384 ATGCTCTTGGCTCCAGAAAGAAACATCCAGCCCTG----- 420
Db 734 -TrpSerCysCys---ArgGlnArgGlnProGlnLeuSerArgGlySerProValProGly 752
QY 421 -----TTAAGTCTCACTCTGCCCCCAGAGTGGAGCTGCTTTTTCAG 467
Db 752 yProArgGlnHisLeuProLeu***MetAlaProArgProTrpLeuTrpAlaSerAla 772
QY 468 ACTGATGAGGCCAGAGTGGTCTGATCCCAACAGACACATATGTAAGGCTCTGGC 527
Db 772 g-Trp-TrpGlyGln----- 776
QY 528 TGACCTATGAGAGGCTCGGCTGACCAAGTACTATCTCAGAGCTGGCTTGCTG 587
Db 776 ----- 776
QY 588 GAGGAGTACTGCACTGGACAGACATGATGTCACCTGGAAACCTCGACAGAAAGCT 647
Db 777 -----AlaProCysThrAlaAlaThrCysCysSerLeu----- 787
QY 648 AACATCCAGACAGACAGATGATGACCAAGCAA----- 680
Db 788 -----ProGlySerLeuGlyLeuAlaTyrArgArgArgPhe 799
  
```

```

QY 681 -----ACGTGCAATTAATCCAAATGTTAAATGATGATTACAGCCCTAGTANGGAC 734
Db 800 ProCysSerCysSerProLeuPheCys-----LeuValPro-ProProTrpAla 816
QY 735 TGCTGGC-----TCTAGTCCAGAAATATGCGGGGATACGCTCTTC 779
Db 816 AlenGlyTrpTrpProThrProAlaAsnSerProGlnSerGlyAlaAlaGlnGluPro 836
QY 780 AACCTG---TGAGCTGTAAGCAAGCTCAGCTAGCTCCCACTGGG----- 825
Db 836 AlaLeuProGluPro***AsnCysLeuArgLeuArgGlyProValGlyArgLeuSer 856
QY 826 -----GCTGTGCCCTCTCT-----CG 842
Db 856 ArgGlyGlnHisArgSerArgAlaHisArgLeuProValHisAsnArgSerGluArg 876
QY 843 GACGTTCCGTGGGACAGCCCATCTG----- 870
Db 876 ySerProProSerSerSerProValIleArgAlaArgProLeuAlaAlaGlyGlnGly 896
QY 871 -----TGTCATATGATGAGAAAG 890
Db 896 yAlaGlySerAlaAlaGlyGluArgLysSerLysGlyProCysSer----- 911
QY 891 TAGCTAAAGCCCTGCTGCTGCTGTCATGCCACAGACAGCGGTGGGCTGCGTG 950
Db 912 -----ArgGlyCysCysCysCysCysCysCysCysCysCysCysCysCysCys 926
QY 951 GGAATTCATCTGAGAGTCTTCTCAGCTTAAGCTTGACAGACAGACTTGCGGGGGA 1010
Db 927 ---Asp-HisArgProPheLeuSerPheGlyGlnAlaGlnAlaLeuThrTrpMetGly 945
QY 1011 TGCTCAGAGATGAGGTGATTTCTGTAACCTGGGAGAGCTATCTCTGACCTCCGACGG 1070
Db 945 yLeuGlnAlaTrpGlnGlySerLysProGlyArgProCysSerLysLeu-----H 962
QY 1071 ACATCTCCAGCGCCAGCCGAGGGGTCA---GGGACAGAGGTGACACCTCAGATGAGCCA 1127
Db 962 lAlaProProProProIleValGlySerGlnSerAlaLysLeuSerCysAla***ArgPro 981
QY 1128 AGACTGGGCTCAGGAGACAG---TGTGTTGAGCCAGACCTG----- 1169
Db 982 AlaLeuGlySerValTrpArgAspLeuCysLeuLeuThrArgGlnIleCysPheCysGln 1001
QY 1170 GGGCGGGGTGGGGCGGGGCGCTTCTGCTCATTTGCTTAAGTAAGAAAGCTCAAGGA 1229
Db 1002 ThrArgPhePheGlyAspProAlaPro-ProSerSerLeuAlaLeuSerGluProGly 1021
QY 1230 GCCAAACCCAGGCTTCCCTCTCGAGTTGAATATCCAGAAATCTTTGACTCTT 1289
Db 1021 nProAlaPro-----ProProSerAla----- 1028
QY 1290 GTTGGTTAATTGTTATTTTGTAAATAAATAAATAAATAGTTAATAAATG 1344
Db 1029 -----ValTrpIlePheIleAspLeuValLeuArgLeuAlaAspArgLeu 1033

RESULT 6
ABG14734
ID ABG14734 standard; Protein; 1040 AA.
XX
AC ABG14734;
XX
DT 18-FEB-2002 (first entry)
XX
DE Novel human diagnostic protein #14725.
XX
KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
XX food supplement; medical imaging; diagnostic; genetic disorder.
XX
OS Homo sapiens.
XX
PN MO200175067-A2.
  
```

PD 11-OCT-2001.
XX
XX 30-MAR-2001; 2001WO-US08631.
PR 31-MAR-2000; 2000US-0540217.
PR 23-AUG-2000; 2000US-0649167.
XX
XX (HYSE-) HYSEQ INC.
PI Dmanac RT, Liu C, Tang YT;
XX WPI; 2001-639362/73.
DR N-PSDB; AAS78921.
XX
XX New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity
XX
XX Claim 20; SEQ ID No 45093; 103bp; English.
XX
XX The invention relates to isolated polynucleotide (I) and
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC and gene mapping, and in recombinant production of (II). The
CC polynucleotides are also used in diagnostics as expressed sequence tags
CC for identifying expressed genes. (I) is useful in gene therapy techniques
CC to restore normal activity of (II) or to treat disease states involving
CC (II). (II) is useful for generating antibodies against it, detecting or
CC quantitating a polypeptide in tissue, as molecular weight markers and as
CC a food supplement. (II) and its binding partners are useful in medical
CC imaging of sites expressing (II). (I) and (II) are useful for treating
CC disorders involving aberrant protein expression or biological activity.
CC The polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC responsible for genetic disorders or other traits to assess biodiversity
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. ABG00010-ABG03077 represent novel human
CC diagnostic amino acid sequences of the invention.
CC Note: The sequence data for this patent did not appear in the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
XX Sequence 1040 AA;
SQ
Alignment Scores:
Pred. No.: 0.000142 Length: 1040
Score: 144.50 Matches: 109
Percent Similarity: 30.77% Conservative: 27
Best Local Similarity: 24.66% Mismatches: 134
Query Match: 5.64% Indels: 172
DB: 22 Gaps: 25
US-09-989-919-15 (1-1397) x ABG14734 (1-1040)
QY 1176 CCGGCCCGAGGCTCTGCTCAACACACCTGCTCCCTGAGCTTGCTGCTGCT 1117
DB 520 ProAlaProGlyCysGlyAla-----SerProGlyProLeuAlaSerAla 535
QY 1116 GAGGTGTGACCTCTGCCCCCTGACCCCTGGGCTGGGCTGGAGTGTCCCTGTGGGAGG 1057
DB 536 -----GlyPro***AlaLeu***GlyThrProSerSer----- 546
QY 1056 TCAGAGATAGCCCTCCAGGTACAGATACACCATCTCTGGAGCATCCCGCCGCAAGTC 997
DB 547 AlaAlaThrSerGlnProGlyThrArgProSerSerThrAlaProProCysGlnAla 566
QY 996 TCCTGTCCAGACTTAAGCTGAGAGAACATCCACGATGATGTGCCCAAGAGCCGCCA 937
DB 567 ProCysPro***ArgProAlaAlaSerProGlyMetGlyAlaProSer----- 583
QY 936 CCGCTGTGTGGCATGTGACGAGCAGACGAGGGGCTTTAGCTACATTCTTCACACTAT 877

DB 584 -----SerCysProAlaAlaSerAlaGly----- 591
QY 876 TGAACACAGTATGAGGCTGCCCGAACCGTCCAGGAGGCGACACCCCGCATGG 817
DB 592 -----ProProHisSerThrAlaGlyAlaHisSerProThrGly 604
QY 816 GGAGACTAGCTGAGCTTGCTTACAGCCACAG-----GGTTGAGA 775
DB 605 ThrProSerProAlaCysAlaAlaGlyProGlnCys***Thr***GlySerGlyTrp--- 623
QY 774 GGCAGTCATACCCCGCATGATTCCTGAGTACAGAGCCAGCATCCCTAGCTAGCTGTA 715
DB 624 ---SerCysTrpPro-----GlyArgSerPheProAla----- 633
QY 714 AACTCAGATTTTAACATTGGCATTTTGCATTTGCTGCTGCTGCTGCTG 655
DB 634 ---SerLeuPheGlySerTrpProAlaCysPro---AlaSerProArgSerSerAlaGln 651
QY 654 GGATGT----- 649
DB 652 GlyCysLeuProLeuAlaAlaArgMetIleValAlaTrpProAlaProAlaAlaGly 671
QY 648 -----TAGCTTGTCTGCAG-----GGGTTCCAGGTGAC 619
DB 672 ProGlnPro***ThrProTrpProCysLeuArgArgGlySer***GlyLeuSerArgSer 691
QY 618 ATGCAGTCTGTCAGTGCAG-----TCACTC-----CCTCCACAG 583
DB 692 TYRGuAlaGlyThrCysArgArgHisSerLeuSerAlaGlyGlyGlyHisProProSer 711
QY 582 GGAAGCCGAGCTGAGGATAGTACAGCTGTCAGCCGAGCCCTGAGATGAGTGCAGCC 523
DB 712 ProAlaProLeuAlaGly-----ValArgSer 720
QY 522 AGCCCTTCATATGATGCTTGTGGATCAGGACCACTTGCTCC----- 475
DB 721 ProProSerAlaThrTrp-----ProGlyProProGlyThrSerSerLeuHis 736
QY 474 -----ATCAGTCTTGAAMAAGACCAGT-----GCCGACTG 442
DB 737 LeuProIleArgSerLeuProArgThrAla**GlyProAlaSerAlaProAlaAspArg 756
QY 441 GGGGGCAGAGGTGAGAGGTAACAGAGGCTGATGTTCTTCTTGGAGGACAGACCATTC 382
DB 757 AlaGlyCysSerGlnProGlyAspGlyHisSer---AlaProLeuThrAlaArgPro--- 774
QY 381 TCCCTAATTGACTGCTTGAAGACACGATATACCTCTCTCTCATCAGCTCAGCACCCCT 322
DB 775 -----AlaAlaProArg 778
QY 321 CTGACTTCGATGGGAGTACTATCTTCAACCTCCAGAGAGGCTCTGCTCTT 262
DB 779 MetProProGlnProGlySer-----SerLeuGlyAlaAlaProAlaLeuHisAla 795
QY 261 GGAAGGGCTTCTCCACAGTGCCTCAGCTGTCAGAGTGTCTGTGTGTCACCAAGAGCC 202
DB 796 SerProAlaGlyProAlaArgGlnPro-----AlaVal-ArgPr 808
QY 201 CGACACTGGGAGTGTCTTCACAGATCCACAGCTCAGCCAGCT----- 156
DB 808 o---ProGlySerProAlaHisArgArgProAlaProGlnProGlyPheProProSerAl 827
QY 155 -----TGAACACTGMAAGAGGACCTCCCGTGGTGGAG 121
DB 827 aProTrpArgProArgLeuSerProArgProValGlyHisGlySerAlaThrProGlySe 847
QY 120 GATGCCAGCAGC-----GGAGCTTCCGG 94
DB 847 rAlaAlaGlnThrTrpSerAlaGlyProArgGlnHisProGlyProGlyThrAlaProG 867
QY 93 GGA 90
DB 867 yGly 868

RESULT 7
ABBI1808
ID ABBI1808 standard; peptide; 1435 AA.
XX
AC ABBI1808;
DT 11-JAN-2002 (first entry)
DE Human BCL9 homologue, SEQ ID NO:2178.
XX
DE Human; cytokine; cell proliferation; cell differentiation; growth factor;
XX haematopoiesis regulation; tissue growth; immunomodulator; activin;
XX inhibin; chemokinesis; chemokinesis; thrombolytic; oncogenesis;
XX proliferation; metastasis; cancer; tumour; haematopoietic disorder;
XX myeloid cell disorder; lymphoid cell disorder; asthma; arthritis;
XX chronic inflammatory condition; proliferative retinopathy;
XX atherosclerosis; coronary heart disease; arterial ischaemia;
XX bone disorder; osteoporosis; vascular growth disorder;
XX tissue regeneration; wound healing; infection; immune disorder;
XX cell culture; drug screening; gene therapy; anti-inflammatory;
XX antihistaminic; antihistaminic; haemostatic; antiarteriosclerotic;
XX cytotactic; osteopatchic; vasotropic; cardiac; virucide; antibacterial;
XX antifungal; vulnery; antitumor.
XX
XX Homo sapiens.
XX
XX WO200157186-A2.
XX
XX 09-AUG-2001.
XX
XX 05-FEB-2001; 2001WO-US03800.
XX
XX 03-FEB-2000; 2000US-0496914.
XX
XX 27-APR-2000; 2000US-0560875.
XX
XX (HYSE-) HYSEQ INC.
XX
XX Tang YT, Liu C, Drmanac RT;
XX
XX WPI, 2001-457740/49.
XX
XX N-PSDB; ABA09052.
XX
XX Human proteins and DNA encoding sequences useful for preventing,
XX treating or ameliorating a medical condition in a mammalian subject
XX e.g. arthritis and cancer -
XX
XX Claim 20; Page 256-257; 1963JP; English.
XX
XX Sequences ABBI0981-ABBI12330 represent 1350 novel human polypeptides, and
XX sequences ABA08225-ABA09574 represent nucleic acids encoding them. The
XX invention also relates to vectors and recombinant host cells comprising a
XX nucleotide of the invention, methods of producing the novel polypeptides,
XX antibodies against the polypeptides, methods of detecting the nucleotides
XX or polypeptides in a sample, and methods of identifying compounds which
XX bind to polypeptides of the invention. Although novel, many of the
XX polypeptides of the invention have homology to known proteins, thereby
XX giving an insight into their probable biological activities, and hence
XX potential therapeutic applications. The polypeptides of the invention may
XX have various activities, including cytokine, cell proliferation or cell
XX differentiation activities; stem cell growth factor activity;
XX haematopoiesis regulatory activity; tissue growth activity;
XX immunomodulatory activity; activity or inhibin-related activities;
XX chemotactic or chemokinetic activities; haemostatic, thrombotic or
XX thrombolytic activities; receptor or ligand activities; or may be
XX involved in oncogenesis, cancer cell proliferation or metastasis.
XX Depending on their biological activities, polypeptides and nucleotides of
XX the invention are useful for preventing, treating or ameliorating medical
XX conditions, e.g., by protein or gene therapy. Such conditions include
XX cancers, haematopoietic disorders (e.g., myeloid or lymphoid cell
XX disorders), chronic inflammatory conditions (e.g., asthma or arthritis),
XX proliferative retinopathy, atherosclerosis, coronary heart disease,
XX arterial ischaemia, bone disorders (e.g., osteoporosis), and abnormal

CC vascular growth. Polypeptides involved with tissue regeneration and
CC repair (or nucleic acids encoding them) may be used to promote wound
CC healing (e.g., of burns, incisions and ulcers), while those with
CC immunomodulatory activities may be used in the treatment of viral,
CC bacterial and fungal infections in addition to immune disorders.
CC Polypeptides with growth factor activity may be used in cell cultures to
CC promote cell growth. For example, such polypeptides may be used to
CC manipulate stem cells in culture to give rise to neuroepithelial cells
CC that can be used to augment or replace cells damaged by illness,
CC autoimmune disease or accidental damage. The polypeptides and nucleotides
CC may also be used in the diagnosis of the above conditions, and in drug
CC screening techniques. The present sequence represents a novel human
CC polypeptide of the invention.
XX
XX
SQ Sequence 1435 AA;
XX
XX
Alignment Scores:
Pred. No.: 0.000471 Length: 1435
Score: 139.50 Matches: 117
Percent Similarity: 30.47% Conservative: 39
Best Local Similarity: 22.85% Mismatches: 157
Query Match: 5.45% Indels: 159
DB: 22 Gaps: 25
US-09-989-919-15 (1-1397) x ABBI1808 (1-1435)
QY 1221 GGCTTCATTGAAGCAATATGAGCAAGGCGCCGCCGCCGCC----- 1174
DB 810 GlyGluHisProGlnIn-----GluTrgYlmetGlyProArgProPheLeuPromet 827
QY 1173 GCGCCAGGTCCTGGCTCA----- 1156
DB 828 SerGlnGlyProGlySerAsnSerGlyLeuArgAsnLeuArgGluProIleGlyProAsp 847
QY 1155 -----ACCACACCTGCTCCTCCAGCCCACTTTG-----GCTCAT 1120
DB 848 GlnArgThrAsnSerArgLeuSerHisMetProPheLeuProLeuAsnProSerSerAsn 867
QY 1119 GCTGAGGTGTGACCTGCGCCCTGACCCCTGGGCTGGGAGTGCCTGTCGGG 1060
DB 868 ProThrSerLeuAsnThrAlaProProValGlnArgGlyLeuGlyArgGlyProLeuAsp 887
QY 1059 AGCTCAGAGATAGCTCCCGCCAGTACAGATCACCACCATCTGGAGCATCCCGCCCA 1000
DB 888 IleSerValAlaGlySer-GlnValHisSer-----ProGlyIleAsnProLeuIly 904
QY 999 GTCTCCTGTCCAGACCTAAGCTGAGAGAACACTCCAC----- 963
DB 904 SerPro-----ThreHisGlnValGlnSerProMetLeuG1 917
QY 962 -----GATGATTTCCTCCACGACGCCCGCCCGCTGCTGTGATGTGA 916
DB 917 YSerProSerGlyAsnLeuIlySerProGlnThrProSerGlnLeu-AlaGlyMetLeuA 937
QY 915 GCGACGACGACGGGGCTTACTTACATTCACACTATTGAACACAGATGCGGGCTCC 856
DB 937 IeGlyProAlaAlaAlaAlaSerIle----- 945
QY 855 CCACGAAACCTGCCAGGAGGGGACACAGCCCGCCAGTGGAGACTAGCTGCTGT 796
DB 946 -----LysSerProProVal-----LeuGlySerAlaAla 956
QY 795 TACAGCCACAGAGGTGTGAGAGGAGCATATACCCCATGATTC-----TGACATAGA 742
DB 956 IAserProValHisLeuIlySerProSerLeuProAlaProSerProGlyTrpThrSers 976
QY 741 GCGACAGATCCCATAGCTAGCGTGTAACTACATTTTAACATTTGGCATTTAGCAG 682
DB 976 eProGluProPro----- 980
QY 681 TTGTCTCGGTGCATCTGTCTGTCTGGAGTTAGCTTTGTCTGAGGGGTTCCAGGT 622
DB 981 -----LeuGlnSerProGlyIleProProA 989

US-09-989-919-15 (1-1397) x AAU97643 (1-738)

[illegible]

Db	291	rProAspLeuProLeuGlyValHisValLeuValLeuThrSerArgProArgCysAla	311
Qy	379	-----CTAATT	372
Db	311	athrLeuArgProAlaArgArgProSerTrpSerSerCysArgAsnSerAlaProGlyPr	311
Qy	371	GACGGCCCTGAAGACAGATATCCCTCTCT	341
Db	331	othrLaserArgCysAlaProProLeuAlaLaProProSerThrThrGlyValLeu	351
Qy	340	----CATACCTCAGACACCCCTCTGACTTCGAT	310
Db	351	uTyrHis-ThrAlaIaIaGlyMetLeuCysAlaAspThrCysAlaGlyProLeuAlaArgA	371
Qy	309	-----GGGGGATCACTATCTATCAACCTCCAGAGAGCTCCGC-----ATCTCC	264
Db	371	laserSer***SerValGlnThrAlaSerSerMetGlyProGlyValCysGlnValAlaPr	391
Qy	263	TTGGAAGGCTTCTCC-----ACTGGCTACCTGTCCAGC	228
Db	391	roGlyArgTrpGlyPro**AlaCysValCysArgAlaAlaIaGly-HisLeuAlaVal	410
Qy	227	TGGTGTGGTGGTGCACCAAGGCCCGACACTGG-----	192
Db	411	MetValGlyTrpThrProSerArgIArgIArgIArgAlaArgCysValValGlyThrThr	430
Qy	191	-----CATGCTCT-----	183
Db	431	AlaArgAlaAlaHisGlyArgAlaLeuSerGlnLeuAlaGlyIaArgGlnAsnMetSerArg	450
Qy	182	---CACAGACATCCACAGCTCAGCCAGCTGAACACTG-----AAAGGAGCAGCTGC	132
Db	451	Phe**GlnLeuProProThr**ProValSerThrLeuProThrThrGlyLeuSerSer	470
Qy	131	CGTGTGGTAGGATGGCCAGCAGCGTAGTCT-----	100
Db	471	HisThrTrpArg**GlySerGlnGlyAlaMetSerTrpLeuGlyArg**AlaSerPro	490
Qy	99	-----TCTGGGGGATGGTCTGTCTGGGATACACTGTACTCG	61
Db	491	LeuThrProProThrProProSerTrpArgMetValValSerSerThrGlnTrpProSer	510
RESULT 9			
AAU32775			
ID	AAU32775 standard; Protein; 384 AA.		
XX			
AC	AAU32775;		
XX			
DT	18-DEC-2001 (first entry)		
XX			
DE	Novel human secreted protein #326.		
XX			
KW	Human; vaccination; gene therapy; nutritional supplement;		
XX	stem cell proliferation; haematopoiesis; nerve tissue regeneration;		
XX	immune suppression; immune stimulation; anti-inflammatory; leukaemia.		
OS	Homo sapiens.		
XX			
PN	WO200179449-A2.		
PD	25-OCT-2001.		
XX			
PF	16-APR-2001; 2001WO-US08656.		
XX			
PR	18-APR-2000; 2000US-055929.		
XX			
PR	26-JAN-2001; 2001US-0770160.		
XX			
PA	(HYSEQ-) HYSEQ INC.		
XX			
PI	Tang YT, Liu C, Drmanac RT;		
XX			
DR	WPI; 2001-611725/70.		
XX			

Nucleic acids encoding a range of human polypeptides, useful in genetic vaccination, testing and therapy -
Claim 20; Page 662; 765pp; English.

The inventors relates to novel human secreted polypeptides. The polypeptides and antibodies to the polypeptides are useful for determining the presence of or predisposition to a disease associated with altered levels of polypeptide. The polypeptides are also useful for identifying agents (agonists and antagonists) that bind to them. Cells expressing the proteins are useful for identifying a therapeutic agent for use in treatment of a pathology related to aberrant expression or physiological interactions of the polypeptide. Vectors comprising the nucleic acids encoding the polypeptides and cells genetically engineered to express them are also useful for producing the proteins. The proteins are useful in genetic vaccination, testing and therapy, and can be used as nutritional supplements. They may be used to increase stem cell proliferation; to regulate haematopoiesis; and in bone, cartilage, tendon and/or nerve tissue growth or regeneration; and immune suppression and/or stimulation; as anti-inflammatory agents; and in treatment of leukemias. AMU2910-AM13304 represent the amino acid sequences of novel human secreted proteins of the invention.

Sequence 384 AA;

Alignment Scores:	
Pred. No.:	0.00666
Score:	125.100
Percent Similarity:	32.58%
Best Local Similarity:	23.76%
Query Match:	4.86%
OB:	22
Length:	384
Matches:	109
Conservative:	39
Mismatches:	158
Indels:	140
Gaps:	21

US-09-989-919-15 (1-1397) X AAU32775 (1-384)

```

QY      1294  CCAACAAGAGTACAAAGATTCTGGATATTCAACTCGAGAGGAGGGAGAAAGCTGGTT 1235
Db      7    ProHnrgIyThrThrTyrLeuProGlyAlaIaGlyAlaValProAspProGlyPheLeu 26
QY      1234  TTG-----GCGCTTTGAGGCTTTCATTGAAAGCAAAATGAGGCGAAGAAAGCCCCG 1184
Db      27    LeuProSerAlaIaLeuSerArgSer---ThrAlaSer***PheSerGlnProPro 45
QY      1183  CCCACACCCCGCCGCCAGGTCCTGGCTCAACACCACTCG---CTCCCTGACCCCGAGTCTT 1127
Db      46    ProValPro-----ProAlaLeuThrValProTyrProArgCys 58
QY      1126  GGCTCATGTGTAGGGGTGCACCTCTGCCCTTGACCCC----- 1090
Db      59    GlyCys***ThSerCysProAlaValProGlnProSerProGlyLeuProArgThrLe 78
QY      1089  -----TGGCGTGGCTGGAGACTGCCCCCTTCGGAGAGTACAGATAGCCTCCCC 1040
Db      78    uValProAlaCysAaArgGlyLeuProLeuSerSerAlaProSerSerProAlaSerArgPr 98
QY      1039  A-----GGTACAGATACACCCACATTCCTGGAGACATCCCGCCCAAGTCTCCTGTCCAGA 986
Db      98    OlyLeuProLeuHHisSerProSerSerTyrSerIleProProGlnGlySerTyrGlyPr 118
QY      985  CCTAAGCTGAGAGAACACTCCACGATGATGTCTCCCAAGCGACGCCCCCAACCGCCTGTGT 926
Db      118  OLeu-----ProHnrgIyThrLeuProLeuPro----- 126
QY      925  GGCATGTGCAGACGACGACGAGGGGCTTTACTACTACATTCTCACACTATGTGAACACAGTg 866
Db      126  ----- 126
QY      865  ATGGGGCTCCACACGGAAACCTGCCAGAGGAGGGGACAGACCCCATGTGGAGACTAGCC 806
Db      127  -TyrGlyLeuProArgGlyLeuGlnSerArgIleu---HisLysPro----- 139
QY      805  TGAAGCTGCTTACAGCCCAAGAGGTGGAGAGGCGTACATACCCCAATGATTCCTGGAGT 746

```

D	b		140	-	-----AlaLeuSerAlaIaIaThrTrpGlnGlyLeuValValAspProSerProHisPr	157
O	y		745	A	GAGCACCACGACTCCCATRGTAGTGGCTGGTAACATCACTTT-----AAAT	698
D	b		157	o	leuLeuAlaPheProLeuLeuSerSerAlaGlnValHisPhe***ProGlyLeuArgSe	177
O	y		697	T	TGGCATTTATTGCAGCTTTGTCCCG-----GTCCATCTGT	662
D	b		177	r	TrpValGlyProPhePheGlyLeuArgLysGlnPro***HisThrGlyLeuSerSeri	197
O	y		661	C	TCCTCGGGATTTTAGCTTTGT-----CTGCAGGGGTTCCAGGT-----	622
D	b		197	u	ThrThrGlnGlyCs-----CysProAlaTrpLeuThrSerTrpProArgLieserGlnse	215
O	y		622	-	-----	622
D	b		215	r	GluArgLysAlaHisPro**CysLeuProLeu**GlnAlaLeuSerProProIleSe	235
O	y		621	-	-----GACATCAGCGTGCCAGAGCAAGCACATCTCCCTCCACAGC	581
D	b		235	r	TYrPheAsnLeuThrValAspLeuArgAsnArgAspTYrHisSerAlaSerSerSe	255
O	y		580	A	AAGCCAGCTGTGAGAGATAGTCACAGCTGTACGCCGACCTCGATAGATAGCAACCAAG	521
D	b		255	r	SerSerSerSerSerSerSerSerSerSerSerSerSerSerArgHrgLeuSerTrpse	275
O	y		520	G	CCCTTCACAT-----ATGTGCTTTGTTGGATCAGGGA---CCACT	482
D	b		275	r	PolySHisHisProLeuLeuProSerPheIleLeuSerTrpProGlnGlyIleProAr	295
O	y		481	T	GGCTCCATCCCACTCTTGAATAAAGACACGCTGCCACTGGGGGGCAGAGGTGAGAGTA	422
D	b		295	g	GlyLeu-----LysAsnProGlyLysThrArgItyGlyLeuGlyTrpTPly	311
O	y		421	A	CAGGGCTGATGTTTCTCTTCTGTGAGGCAAGACCATTTCCCTAAATTGACTGCCCTTG	362
D	b		311	e	LysGlyPheLeuAlaAsnTrpLeuSerLysSerPro-----	322
O	y		361	A	AGACAGTATACCTCTCTCTCATCATCACTCAGCACCCCTCTGACT--TCCGATGGGGG	305
D	b		324	-	-----ArgArgLysLysGlyProSerPheProProLeuValLeu**PheGlnG	340
O	y		304	A	TCTATCTATCATCAACCTCCAGAGAGCTCTGCCATTCCTCTGGAAGGCTTCTCCAC	245
D	b		340	y	ProArgAlaLeuProValProHrgAsnProHisProLysLeuGlyAlaSerSerGln-	355
O	y		244	T	GGCTCACTGTCCAGATGGTGTGGTGTGACCAAAGGCCCACTGGGAGTGGCT	185
D	b		360	-	-----SerPro-SerProTrpTPlyGlyLysArgProLysProLysSerProGlyLysnglyT	378
O	y		184	C	T	183
D	b		378	h	r	378

RESULT 10
 AAR45945 ID AAR45945 standard; Protein; 1239 AA.
 AAR45945;
 08-AUG-1994 (first entry)
 Glutamic acid receptor.
 Glutamic acid receptor; synaptic signal translaion; diagnosis;
 brain disease; nerve signal; gene therapy; NMDA; cerebellum; ICR;
 mouse; nerve cell necrosis.
 Mus musculus.

FT	Key	Region	Location/Qualifiers
FT	Key	Region	location/Qualifiers
FT	Key	Region	361..369
FT	Key	Region	/note= "this line is omitted in the specification"

FT Region 622..630
 FT /note="This line is omitted in the specification"
 FT Region 883..891
 FT /note="This line is omitted in the specification"
 XX JP06014783-A.
 PD 25-JAN-1994.
 PF 30-JUN-1992; 92JP-0173155.
 PR 30-JUN-1992; 92JP-0173155.
 XX (MITU) MITSUBISHI KASEI CORP.
 PA
 DR WPI: 1994-061478/08.
 XX N-PSDB; AA056916.
 XX
 PT New glutamic acid receptor and gene - for use in analysis of
 PS synaptic signal transduction, and diagnosis of brain disease
 CC Claim 1, Page 25-31; 35pp; Japanese.
 CC The sequence shows a glutamic acid receptor. The receptor is
 CC useful for the analysis of nerve signal transduction; within the
 CC synapse, expression of synapse plasticity, nerve cell necrosis,
 CC brain structure and brain disease. It can also be used in gene
 CC therapy.
 XX
 SQ Sequence 1239 AA;
 Alignment Scores:
 Pred. No.: 0.0128 Length: 1239
 Score: 124.00 Matches: 94
 Percent Similarity: 30.63% Conservative: 23
 Best Local Similarity: 24.61% Mismatches: 112
 Query Match: 4.84% Indels: 153
 DB: Gaps: 24
 US-09-989-919-15 (1-1397) x AAR45945 (1-1239)
 QY 1196 AGAAGAGCCCGGCCCCCGCCCGCCAGCTCTGCTCAACACACACTGCTCCCTGA 1137
 DB 927 ArgArgAlaProAlaPro-----ThrThrSerGlyPro----- 937
 QY 1136 CCCCAGCTTGGCTCATGCTGAGGTGCACTCTGCCCCCTGACCCCTGGG----- 1086
 DB 938 -----ArgSerCysThrProGlyPro-----ProGlyGlnProSer 949
 QY 1085 ---CTGGCTGGAGAGTGTCCCTGTGTGGGAGGTGAGATAGCCTCCCGAGTACAGAT 1029
 DB 950 ProSerGlyTyrPargProPro-----GlyGlyGlyArg-----ThrProAlaAlaArgArg 966
 QY 1028 CACCCACATCTGAGAGATCCCGCCCAAGTCTCTGTCAGACCTAAGCTGAGAGACA 969
 DB 967 AlaProGlnProProAlaArgPro-----GlyProAlaGln----- 978
 QY 968 CTCACAGATGATTTGTCACAGAGCCCGCCAGCCGCTGTGGCATGTGCAGACGACG 909
 DB 979 -----GlyArgLeuSerProThrCysProGlnHis----- 988
 QY 908 CAGCAGAGGCTTAAAGTACATCTTCACACTATTAACACAGTATGAGGGGCTGCCACGGA 849
 DB 989 -----ProAlaGlyThrLeuGlyMetArgGlyGlnCysGluSer 1002
 QY 848 ACCGTCACAGAGGAGGAGCAGCCCGCAGTGGGAGACTAGCTG-----AGCTTG 798
 DB 1003 GlyIleArgAspArgThrSerArgProProGluArgArgAlaLeuProGluArgSerLeu 1022
 QY 797 CTTACAGCCCAAC-----AGGTTGAGAGAGGACATC 768
 DB 1023 LeuHisIleHisCysHisItyrSerSerPheProArgAlaGluArgSerGlyArgProPhe 1042

QY 767 ATACCCCATGATTCCTGAGTACAGACCCAGA-----GTCCCATAGTACGCC 720
 DB 1043 LeuPro-----LeuPheProGluProProGluProAspAspLeuProLeuGly 1059
 QY 719 TGGTAACATCACATTTAACATTGGCATTAATTCACGCTTGCTCCGTGCATCTGTCT 660
 DB 1060 ---ProGluGlnLeuAlaArgArgGluAlaLeuLeuArgAlaAlaTrpAla----- 1075
 QY 659 GTCTGGAGATTAGCTTGTCTGACAGGGGTCCAGG---TGACATGAGTGTGCAGT 603
 DB 1076 -----ArgGly-ProArgProArgHisAlaSerLeuProSe 1087
 QY 602 GCAAGTC-----ACTCCCTCCACAGGCAAGCCACG---TGCTGAGATAGTC 558
 DB 1087 rSerValAlaGluAlaPheThrArgSerAsnProLeuProAlaArgCysThrGlyHisAl 1107
 QY 557 AGCTGTACAGCCGAGCCCTCAGATAGGTACAGCAGAGCCCTTCACATATGTGTCTGT 498
 DB 1107 ACysAlaCysProCysProGln----- 1114
 QY 497 GGGATCAGGAGCACCTTGCTCCATCCAGCTTGAAAGACAGCTGCCGA----- 446
 DB 1115 -----SerArgProSerCysArgHisValAl 1123
 QY 445 -----CCTGGGGGAGAGGTAGAGTAACAGAGGCTGAGT 408
 DB 1123 aglnThrGlnSerLeuArgLeuProSerItyrArgGluAlaCysValGluGlyValProAl 1143
 QY 407 TTCTCTTTCTGGAGGACAGACCATCTCCCTAATTGACTGCTTGAAGACAGTATACC 348
 DB 1143 aglyValAlaAla-----ThrTrpGlnPr 1151
 QY 347 TCTCTCTCAT-----CACCTCAGACCCCTCTGACTTCGANGGGGATCACTAT 297
 DB 1151 oArgGlnHisValCysLeuHisThrHisThrHisLeuProPheCysItyrGlyThrValCys 1171
 QY 296 CTATCAACCTCCAGAGAGGTCTGCACTCTCCCTGGAAGGGCTTCCCATGCTGCAC 237
 DB 1171 sArgHisProProProCysSerSerHisSerProTrp----- 1183
 QY 236 CTGTCCAGGTGTGTGTTGTT-----GACCACAAAGGCCGACACTGGG 192
 DB 1184 -----LeuIleGlyThrTrpGluProProSerHisArgGlyArgThrLeuGln 1199
 QY 191 CARG 188
 DB 1199 yLeu 1200
 RESULT 11
 ID ABG03717 standard; Protein; 505 AA.
 XX ABG03717;
 AC
 DT 13-FEB-2002 (first entry)
 XX
 DE Novel human diagnostic protein #3708.
 KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
 KW food supplement; medical imaging; diagnostic; genetic disorder.
 OS Homo sapiens.
 XX
 PN WO200175067-A2.
 PD 11-OCT-2001.
 PF 30-MAR-2001; 2001WO-US08631.
 PR 31-MAR-2000; 2000US-0540217.
 PR 23-AUG-2000; 2000US-0649167.
 XX (HYSE-) HYSEQ INC.

XX 13-FEB-2002 (first entry)
 XX Novel human diagnostic protein #3524.
 DE Human; chromosome mapping; gene mapping; gene therapy; forensic;
 XX Food supplement; medical imaging; diagnostic; genetic disorder.
 KW Homo sapiens.
 OS WO200175067-A2.
 PN 11-OCT-2001.
 PD 30-MAR-2001; 2001WO-US08631.
 PF 31-MAR-2000; 2000US-0540217.
 PR 23-AUG-2000; 2000US-0649167.
 XX (HYSE-) HYSE INC.
 XX Dmanac RT, Liu C, Tang YT.
 PI WPI, 2001-639362/73.
 DR N-PSDB; AAS67720.
 XX New isolated polynucleotide and encoded polypeptides, useful in
 PT diagnostics, forensics, gene mapping, identification of mutations
 PT biologically for genetic disorders or other traits and to assess
 PT biodiversity.
 XX Claim 20; SEQ ID No 33892; 103bp; English.
 XX The invention relates to isolated polynucleotide (I) and
 CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
 CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
 CC and gene mapping, and in recombinant production of (II). The
 CC polynucleotides are also used in diagnostics as expressed sequence tags
 CC for identifying expressed genes. (I) is useful in gene therapy techniques
 CC to restore normal activity of (II) or to treat disease states involving
 CC (II). (II) is useful for generating antibodies against it, detecting or
 CC quantitating a polypeptide in tissue, as molecular weight markers and as
 CC a food supplement. (II) and its binding partners are useful in medical
 CC imaging of sites expressing (II). (I) and (II) are useful for treating
 CC disorders involving aberrant protein expression or biological activity.
 CC The polypeptide and polynucleotide sequences have applications in
 CC diagnostics, forensics, gene mapping, identification of mutations
 CC responsible for genetic disorders or other traits to assess biodiversity
 CC and to produce other types of data and products dependent on DNA and
 CC amino acid sequences. ABG00010-ABG30377 represent novel human
 CC diagnostic amino acid sequences of the invention.
 CC Note: The sequence data for this patent did not appear in the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences.
 CC XX
 SQ Sequence 900 AA;
 Alignment Scores:
 Pred. No.: 0.0157 Length: 900
 Score: 122.50 Matches: 96
 Percent Similarity: 31.04% Conservative: 44
 Best Local Similarity: 21.29% Mismatches: 156
 Query Match: 4.79% Indels: 156
 DB: 22 Gaps: 19
 US-09-989-919-15 (1-1397) x ABG03533 (1-900)
 QY 1194 AAGAGCCGCGCCGAGGCTGCTCAAAACACACACTGCTCCTGACC 1135
 Db 489 ArgGlyProGlyProThrAlaAlaGlyPro----- 499
 QY 1134 CCAAGCTTGAGCTATGAGGTGACCTTGGCCCTGACCCCTGGGCTGAGGA 1075
 Db 1134 CCAAGCTTGAGCTATGAGGTGACCTTGGCCCTGACCCCTGGGCTGAGGA 1075

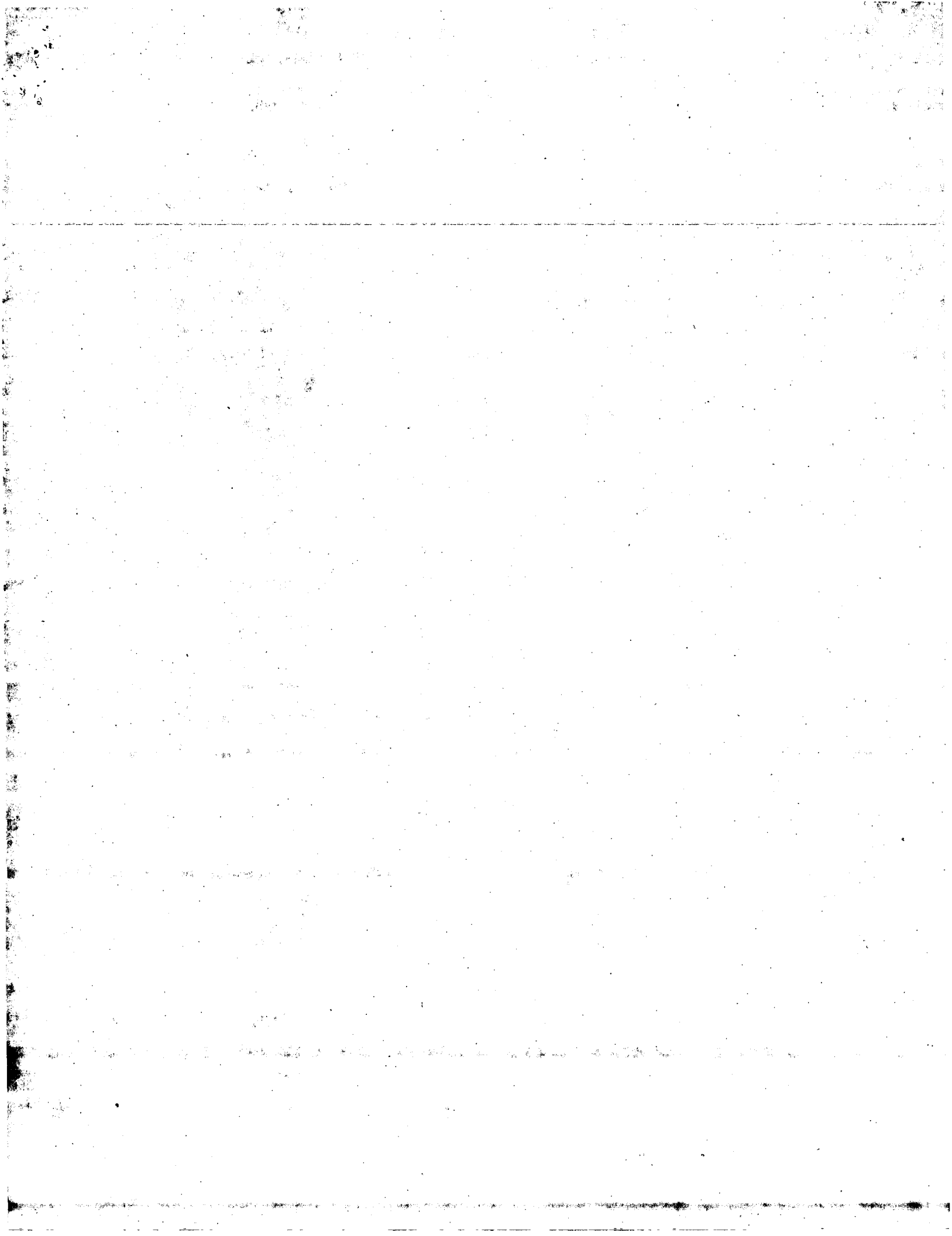
Db 500 -----AlaAlaProGlyArgGlyArgAlaGly 509
 QY 1074 GTGTCCCTGTGTGGAGGTGAGATAGCTCCGAGTACAGAACATCCTCTG 1015
 Db 510 SerAlaAlaArgGlyProGlyProAlaAlaProAlaArgGlyProAlaAla 529
 QY 1014 AGCATCCCCCGCCAGAGTCTCTGTCAGACTGAGAGAACATCAGCA 961
 Db 530 ArgLeuProArgThr-----ArgGlyProProArgProAla 541
 QY 960 -----TGATTGTCCCGACGAGCCCGCCCTGCTGTGCAATGACAGA 913
 Db 542 ProSerGluProThrProLeuProSerArgProGlySerGlyLysArgGlyAlaSer 561
 QY 912 GCAGAGCAGAGGGGCTTACTTACTTCACTATTGAACAGATGAGGGGCTGCCA 853
 Db 562 AlaProProGly-----GlyCysPro 568
 QY 852 CGGAACCTGTCCAGAGGAGGAGCAGAGCCCGAGTGGAGACTGAGCTGCTTAC 793
 Db 569 -----SerArgSerGlyLeuSerProThrLeuGlnAspProGlyProHisSerGly 585
 QY 792 AGCCACAGAGGTTG-----AGAGCAGTCATACCCCATGATTCT 751
 Db 586 CysProGluGlyPheProArgLeuAlaGlyGlyLysArgSerGlyProGluMetGly 605
 QY 750 GGACTAGAGCAGCAGAGTCCAGTACGAGTGTAACTCATTTTAACTTGGAT 691
 Db 606 ThrValGlySerAsnAspValAlaGlyAsnArgGlyLysThr---GlySerSerTrp--- 624
 QY 690 TATTGACGCTTGTCTGTGTCATCTGTCTGTGGATGTAGCTTGTGTGACGGG 631
 Db 624 ----- 624
 QY 630 TTCCAGAGTGCATGATGCTGCTGTCAGTCAAGTACTCTTCACAGCAAGCCAGCT 571
 Db 625 -----ValGlnArgThrTrpAlaGlyGluLeuProProGlySerMetGly 640
 QY 570 GGTGAGGATGAGTGTGTCAGCGAGCTCAGTACAGTACAGCCAGAG----- 521
 Db 641 Ala---GlyThrGluTrpGlyGlyGlyProGluAsnProLysProAlaArgGlyAlaC 660
 QY 520 -----GCTTTCATATATGTGTCTTGTGGATCAGAGCACTTGTCTCATCCAG 469
 Db 660 yGlyCysSerPheThrTrpGlyAlaGlnAlaGlyLysValProGluTrpGlyGlnSerS 680
 QY 468 TCTGAAAAGACACAGC-----TGCCAGCTGGGGGGCAGA 433
 Db 680 eAlaAlaGlySerLysGlySerGlnAlaProSerAsnLeuGlyGlnAlaGlyProGlnAla 700
 QY 432 GGTGAGAGTAAACAGGCTGATGTTCTTCTTGTGAGGCAAGACATTCCTCAAT 373
 Db 700 IaglyGlnGlnGlyGlnArgSerMetProLysArgGlyLysCysThr-----C 716
 QY 372 TGACTGCTTGAAGACACAGTATACCTCTCTCATCATCCTCAGCAAGCCCTGACTTC 313
 Db 716 ySerGlnCysHisValLysLeuGlnLeuGlnAlaLeuValGlyProGlySerThrS 736
 QY 312 GATGGGGATCACTATCTATCAACCTCAGAGAGGCTCTGCCATCTCCTTGAGAGGGCT 253
 Db 736 IYTrp-----LysCysHis 741
 QY 252 TCTCCACAGTGGCTACCTGTCCAGG-----GTCGTGTGGTGA----- 213
 Db 742 -----HisGlnSerSerValProGlyHisSerAlaProProGlyLeuIleGlyLysSerS 760
 QY 212 -----CCACAAGGCGCCGACACTGGG-----CATGGC----- 186
 Db 760 IYProIleAlaGlyProArgAlaProGlyCysGlyLeuArgLysGlnGlyLysGlyProG 780
 QY 185 -----TTCACAGACATCCACAGCTCAGCCAG 158
 Db 780 IYLeu---AlaThrPheGlyGlyAlaProLeuSerGluArgValPro---GlySerProP 800

QY 157 GTTGAACACTGAAGAGCAGCTCCCTGTGTGTAGATGACAGCGGTACTTTC 98
 Db 800 rcaalallerproysgllyhlsprothclproserSerProgllyserTTP 820
 QY 97 CTGGG-----GGATGTCTGTCTGTGGATACCTGTACTCGTGTGT 50
 Db 820 rogllyProArgrgluSerGlyTTP-----ValSerValThrGly-----L 833
 QY 49 TGCCGTGAGATTCTGCAGATACTGCCCCG 21
 Db 833 euProGlnThrAlaAlaGlyProArgrArg 842
 RESULT 15
 AAY44789
 ID AAY44789 standard; Protein; 343 AA.
 AC AAY44789;
 AC AAY44789;
 DT 04-MAY-2000 (first entry)
 XX
 DE "Corn Adenylsulphate kinase-2."
 XX
 KM Adenylsulphate kinase; Adenosine-5'-phosphosulphate kinase; APS kinase;
 KM 3'-phospho-adenosine-5'-phosphosulphate; PAPS; sulphate assimilation;
 KM corn; clone p0016.ctscj40rb; transgenic plant; screen; antibody.
 OS Zea mays.
 PN W0200004165-A1.
 XX
 PD 27-JAN-2000.
 XX
 PF 13-JUL-1999; 99MO-US15809.
 XX
 PR 14-JUL-1998; 98US-0092833.
 XX
 PA (DUPO) DU PONT DE NEMOURS & CO E I.
 XX
 PI Falco SC, Allen SM, Anderson SL;
 XX
 DR WPI; 2000-182430/16.
 DR N-PSDB; AAZ50160.
 XX
 PT New nucleic acid molecule and chimeric gene encoding an adenosine-5'
 PT phosphosulphate kinase, useful for altering expression of sulfate
 PT assimilation protein in plants
 XX
 PS Claim 6; Page 31-32; 42pp; English.
 XX
 CC The present sequence is the corn adenylsulphate kinase (APS kinase),
 CC also known as adenosine-5' phosphosulphate kinase. This is obtained from
 CC clone p0016.ctscj40rb, derived from corn pooled tassels shoots, p0016
 CC cDNA library. This has 528 sequence identity to Cathartus roseus APS
 CC kinase. APS kinase is a sulphate assimilation protein, that catalyses
 CC the conversion of adenosine-5' phosphosulphate (APS) to 3'-phospho-
 CC adenosine-5' phosphosulphate (PAPS). The nucleotide sequence is used as
 CC probes and primers to identify, obtain and synthesize sulphate
 CC assimilation proteins from other plants. It is also used to produce
 CC transgenic plants, that are useful for altering the expression levels of
 CC a sulphate assimilation protein. The APS kinase peptides are useful for
 CC producing antibodies, that are used to screen and isolate cDNA clones.
 XX
 SQ Sequence 343 AA;
 Alignment Scores:
 Pred. No.: 0.0123 Length: 343
 Score: 122.00 Matches: 88
 Percent Similarity: 35.52% Conservative: 42
 Best Local Similarity: 24.04% Mismatches: 118
 Query Match: 4.77% Indels: 118
 DB: 21 Gaps: 19

US-09-989-919-15 (1-1397) x AAY44789 (1-343)
 QY 1182 CCCACCCCGCCCGAGGCTCTGGCTCA-----AACACACCTGCTCC 1141
 Db 20 ProSerProAlaProGlyProAlaSerGlnGlyAlaArgGlnGlyAlaThrLeuLeuSer 39
 QY 1140 CTGACCCCGAGCTGGGCTGAGGTGAGGTGACACTCTGCGCCCTGACCCCTGGGCTGGC 1081
 Db 40 ProThrProThrLeuAla-ValIleLeuValAsnPro----- 51
 QY 1080 CTGGAGAGTGCCCTGTGGAGAGGTGACAGATATACCTCCAGATACAGATACACCA 1021
 Db 52 -----GlnArgAlaProProValLeu----- 58
 QY 1020 TCTTGAGACATCCCGCCCAAGTCTCTGTGTCCAGACCTAACCTGAGAGACATCCACGA 961
 Db 59 -ProGlyLeuThrProSerAspAlaProLeuProAla-----LeuValIleHis-- 74
 QY 960 TGGATTGTCGCCAGCAGCCCGCCACCGCTCTGTGAGTGTGACAGCAGCAGC----- 908
 Db 75 -GlyLeuThrProAlaGlySerSerHis-SerSerAlaGlyLeuAlaSerSerGlyArg 94
 QY 907 -----AGCAGGGGCTTTAGCTACATTTCTTCACTATTGAACACAGTATGGGCG 859
 Db 94 rggluglygluglyarggly-AlaArgThrHisCysHisArgglyileglyArgTTP--- 112
 QY 858 TGCCACCGGACCCGTCGCCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 799
 Db 113 --ValArgArgArgArgArgArgArgArgArgArgArgArgArgArgArgArgArg 126
 QY 798 GCTTACAGCCCGCAGAGGTTGGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 739
 Db 127 ProHisSerProVal-----LysGlnLysProValMetSerAsnIleGlyLys 142
 QY 738 AGCAGTCCCATAGCTAGGCTGTAACTCACTTTTAACTTTGACATTTATGTCACAGTTT 679
 Db 143 SerThrAsnIle-----LeuThrIleAsnLysLeuIle 153
 QY 678 GTTCGTGTACATCTGTCTGTCTG-----GGATGTACTTGTGTGAGGAGG 631
 Db 154 GlyGlnSerAspArgGlnLeuLeuLeuGlnGlyGlyCysValValIleThrIleThrGly 173
 QY 630 TTCACAGGTGACATGACG-----TGTCGCCAGTGCAGTCACTCCCTCCACAGGCA 580
 Db 174 LeuSerGlySerGlyLysSerThrLeuAlaCys-AlaLeuSerArgGlyLeuHisCysArg 193
 QY 579 AGCCACGCTGCTG-----AGCATGT 559
 Db 193 ggllyHisLeuThrIleValLeuAspGlyAspAsnLeuArgHisGlyLeuAsnArgAsp 213
 QY 558 CAGCTGTGACGCGAGCCCTCAGATAGTCA-----GC 526
 Db 213 uSerPheLysAlaGlu-----AspArgAlaGluAsnIleArgArgValGlyValAla 231
 QY 525 CAGAGGCTTCACATATGTGTCTGTGGATCAGAGGACCACTTGCTCCATCCAGTGT 466
 Db 231 AluLysLeuPheAlaAspAlaGlyValIleCysIleAlaSerLeuIleSerProTyrArgArg 251
 QY 465 TGAATAAGAC-----CA 454
 Db 251 gAspArgAspAlaCysArgAlaLeuLeuProHisSerAsnPheIleGluValPheIleAs 271
 QY 453 GCTGCCGACCTGGGGGAGGAGGTGAGAGGTGAGAGGCTGATGTTCTCTTCTGAG 394
 Db 271 pLeuProLeuLysIleCysGluAlaArgAspProLys--GlyLeuTyrIleLeuAlaArg 290
 QY 393 GCAGAAGCATTTCTCCTAATTGACTGCTTGAAGACAGATATCTCTCTCATACAC 334
 Db 290 gThrGlyLysIleLysGlyPheThrGlyLeaAspAspProTyr----- 304
 QY 333 TCAGACACCCCTCTGACTCCAGATGGGAGATACATATGATACCTCCAGAGAGTCT 274
 Db 305 -GluProPheIle-----AsnIleGlnIleValIleLysMetLysAspIleGlu 321

OY 273 GCCATCTCCC 264
|||
Db 321 sProSerPro 324

Search completed: June 24, 2003, 13:52:29
Job time : 103.5 secs



GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on: June 24, 2003, 13:49:32 / Search time 23 Seconds
(without alignments)
3574.245 Million cell updates/sec

Title: US-09-989-919-15

Perfect score: 2527

Sequence: 1 99tgcgtcaccctacccgga.....aaaaaaaaaaagcggtc 1397

Scoring table: BLOSUM62
Xgapop 10.0, Xgapext 0.5
Ygapop 10.0, Ygapext 0.5
Fgapop 6.0, Fgapext 7.0
Delop 6.0, Delext 7.0

Searched: 262574 seqs, 29422922 residues

Total number of hits satisfying chosen parameters: 525148

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame_plus_n2p_model -DEV=xld
-O=/cg2_1/USFTO_spool/US0998919/runat_24062003_102048_19220/app_query.fasta_1.1543
-DB=Issued_Patents_AA -OPMT=fastan -SUFTX=rai -MINMATCH=0.1 -DOOPCL=0
-DOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blomsu62 -TRANS=human40.cdi
-LIST=45 -DOCALLIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US0998919 @CGN 1.1 28 @runat_24062003_102048_19220 -NCPU=6 -ICPU=3
-NO MAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSFBLOCK=100 -LONELC
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREDS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database:

- 1: /cg2_6/prodata/1/iaa/5A_COMB.pep.*
- 2: /cg2_6/prodata/1/iaa/5B_COMB.pep.*
- 3: /cg2_6/prodata/1/iaa/5A_COMB.pep.*
- 4: /cg2_6/prodata/1/iaa/5B_COMB.pep.*
- 5: /cg2_6/prodata/1/iaa/PCTUS_COMB.pep.*
- 6: /cg2_6/prodata/1/iaa/backfile1.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length DB	ID	Description
C 1	124	4.8	1239	1	US-08-026-138E-3 Sequence 3, Appl1
C 2	120	4.7	1964	4	US-09-467-997-1 Sequence 1, Appl1
C 3	119	4.6	1063	1	US-08-093-453B-3 Sequence 3, Appl1
C 4	118	4.6	897	1	US-07-860-389-2 Sequence 2, Appl1
C 5	117	4.6	1185	4	US-09-041-886-23 Sequence 23, Appl1
C 6	116	4.5	433	4	US-09-046-158A-2 Sequence 2, Appl1
C 7	115	4.5	907	3	US-08-783-774-2 Sequence 2, Appl1
C 8	115	4.5	907	5	US-09-328-599A-1 Sequence 1, Appl1
C 9	115	4.5	907	5	PCT-US95-04611A-19 Sequence 19, Appl1
C 10	115	4.5	1063	1	US-08-127-499A-8 Sequence 8, Appl1
C 11	115	4.5	1063	1	US-08-482-847-8 Sequence 8, Appl1
C 12	112.5	4.4	376	4	US-08-874-569B-20 Sequence 20, Appl1

C 13	112	4.4	992	1	US-08-127-499A-1 Sequence 1, Appl1
C 14	112	4.4	992	1	US-08-482-847-1 Sequence 2, Appl1
C 15	109.5	4.3	1298	2	US-08-690-473-2 Sequence 2, Appl1
C 16	109.5	4.3	1298	4	US-09-259-821A-2 Sequence 2, Appl1
C 17	109.5	4.3	1298	4	US-08-843-655-2 Sequence 2, Appl1
C 18	109	4.3	1130	4	US-09-442-100-4 Sequence 4, Appl1
C 19	106	4.1	878	4	US-09-556-706B-2 Sequence 2, Appl1
C 20	105	4.1	557	4	US-08-927-213-139 Sequence 139, App
C 21	105	4.1	962	4	US-09-442-100-6 Sequence 6, Appl1
C 22	100.5	3.9	708	1	US-08-396-479B-8 Sequence 8, Appl1
C 23	100.5	3.9	708	1	US-08-818-823-8 Sequence 8, Appl1
C 24	100.5	3.9	739	1	US-08-396-479B-10 Sequence 10, Appl1
C 25	100.5	3.9	739	1	US-08-818-823-10 Sequence 10, Appl1
C 26	100.5	3.9	1068	1	US-08-396-479B-12 Sequence 12, Appl1
C 27	100.5	3.9	1068	1	US-08-818-823-12 Sequence 12, Appl1
C 28	100.5	3.9	1075	5	PCT-US94-07297-41 Sequence 41, Appl1
C 29	100.5	3.9	1214	2	US-08-231-193A-54 Sequence 54, Appl1
C 30	100.5	3.9	1214	2	US-08-486-273A-54 Sequence 54, Appl1
C 31	100.5	3.9	1214	3	US-08-480-474-54 Sequence 54, Appl1
C 32	100.5	3.9	1214	3	US-08-940-086A-54 Sequence 54, Appl1
C 33	100.5	3.9	1214	4	US-08-940-035A-54 Sequence 54, Appl1
C 34	100.5	3.9	1214	4	US-08-935-105A-54 Sequence 54, Appl1
C 35	100.5	3.9	1214	4	US-09-648-797-54 Sequence 54, Appl1
C 36	100.5	3.9	1219	2	US-08-231-193A-50 Sequence 50, Appl1
C 37	100.5	3.9	1219	2	US-08-486-273A-50 Sequence 50, Appl1
C 38	100.5	3.9	1219	3	US-08-480-474-50 Sequence 50, Appl1
C 39	100.5	3.9	1219	3	US-08-940-086A-50 Sequence 50, Appl1
C 40	100.5	3.9	1219	4	US-08-940-035A-50 Sequence 50, Appl1
C 41	100.5	3.9	1219	4	US-08-935-105A-50 Sequence 50, Appl1
C 42	100.5	3.9	1219	4	US-09-648-797-50 Sequence 50, Appl1
C 43	100.5	3.9	1231	2	US-08-231-193A-48 Sequence 48, Appl1
C 44	100.5	3.9	1231	2	US-08-486-273A-48 Sequence 48, Appl1
C 45	100.5	3.9	1231	3	US-08-480-474-48 Sequence 48, Appl1

ALIGNMENTS

RESULT 1
US-08-026-138E-3
Sequence 3, Application US/08026138E
Patent No. 5502166
GENERAL INFORMATION:
APPLICANT: Masayoshi MISHINA
TITLE OF INVENTION: NOVEL PROTEINS AND GENES CODING THE SAME
NUMBER OF SEQUENCES: 19
CORRESPONDENCE ADDRESS:
ADDRESSEE: Nishionata Residence 1-107
STREET: 5214, Nishionata-machi
CITY: Niigata-shi
STATE: Niigata-ken
COUNTRY: JAPAN
ZIP: 951
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3.50 inch, 1.44 MB storage
COMPUTER: IBM Compatible
OPERATING SYSTEM: MS-DOS v.5
SOFTWARE: Word Perfect 5.1
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/026,138E
FILING DATE: 26-FEB-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 39563/1992
FILING DATE: 26-FEB-1992
APPLICATION NUMBER: JP 173155/1992
FILING DATE: 30-JUN-1992
APPLICATION NUMBER: JP 215017/1992
FILING DATE: 12-AUG-1992
APPLICATION NUMBER: JP 303878/1992
FILING DATE: 13-NOV-1992
ATTORNEY/AGENT INFORMATION:
NAME: Hamburg, C. Bruce
REGISTRATION NUMBER: 22,389
REFERENCE/DOCKET NUMBER: F-4551

TELECOMMUNICATION INFORMATION:
 TELEPHONE: (212) 986-2340
 TELEFAX: (212) 953-7733
 INFORMATION FOR SEQ ID NO: 3:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 1239 amino acids
 TYPE: amino acid
 STRANDEDNESS: single strand
 TOPOLOGY: linear
 MOLECULE TYPE: protein
 ORIGINAL SOURCE:
 ORGANISM: mouse
 TISSUE TYPE: brain
 PUBLICATION INFORMATION:
 AUTHORS: Masayoshi MISHINA
 TITLE: NOVEL PROTEINS AND GENES CODING THE SAME
 RELEVANT RESIDUES IN SEQ ID NO: 3: FROM 1 to 1239
 US-08-026-138E-3

Alignment Scores:

Pred. No.:	0.000518	Length:	1239
Score:	124.00	Matches:	94
Percent Similarity:	30.63%	Conservative:	23
Best Local Similarity:	24.61%	Mismatches:	112
Query Match:	4.84%	Indels:	153
DB:	1	Gaps:	24

US-09-989-919-15 (1-1397) x US-08-026-138E-3 (1-1239)

```

QY 1196 AGAAGAGCCGCGGCGCCGCGCCGAGTCTGCTCAACACACAGCTGCTCCCTGA 1137
DB 927 ArgArgAlaProAlaPro-----ThrThrSerGlyPro----- 937
QY 1136 CCCAGCTTGAGCTGAGTGTGACCTCTGCCCCCTGACCCCTGGG----- 1086
DB 938 -----ArgSerCysThrProGlyPro-----ProGlyGlnProSer 949
QY 1085 ---CTGGCTGGAGAGTGTCTCCCTGTGCGGAGTCAAGATAGCTCCCGAGTACAGAT 1029
DB 950 ProSerGlyTrpArgProPro-----GlyGlyGlyArg---ThrProLeuAlaArgArg 966
QY 1028 CACCCACATCTGGAGATGCCCGGCAAGTCTCTCCAGACCTAAGCTAGAGAGACA 969
DB 967 AlaProGlnProProAlaArgPro-----GlyProAlaGln----- 978
QY 968 CTCACAGATGATGTGCTCCGACGACGCCCGCCGCTGTGTGGCATGTGCAGCAGCAG 909
DB 979 -----GlyArgLeuSerProThrCysProGlnHis----- 988
QY 908 CAGCAGGGGGCTTAAAGTACATTCTCACATTAATTAACAGATGATGGGGCTGCCACGGA 849
DB 989 -----ProAlaGlyThrLeuGlyMetArgGlyGlyGlnCysGlnSer 1002
QY 848 ACCGTCCAGGAGAGGGGACAGCCCGCAGTGGGAGACTAGGCTG-----AGCTTG 798
DB 1003 GlyIleArgAspArgGlnThrSerArgProProGlnArgArgAlaLeuProGlnArgSerLeu 1022
QY 797 CTACAGCCGAC-----AGGGTTGAGAGAGCAGTC 768
DB 1023 LeuHisAlaHisCysHisIleSerSerPheProArgAlaGlnArgSerGlyArgProPhe 1042
QY 767 ATACCCCAATGATCTCTGAGCTAGAGCCAGCA-----GTCCCATAGAGCTAGGC 720
DB 1043 LeuPro-----LeuPheProGlnProProGlnProAspAspLeuProLeuLeuGly 1059
QY 719 TGGTAACTACATTTTAACTTTGGCATTTTGCAGCTTGTCTGTGCTGACATCTGTCT 660
DB 1060 ---ProGlnGlnLeuAlaArgArgGlnAlaLeuLeuArgAlaAlaIlePala----- 1075
QY 659 GTCTGGAGATTTACTTGTCTGTCGAGGGGTTCCAG---TGACATGACAGTGTCTGCCAGT 603
DB 1076 -----ArgGly-ProArgProArgHisAlaSerLeuProSe 1087
  
```

```

QY 602 GCAAGTC-----ACTCCCTCCAGCAGGACCCGAC---TGCTAGGATAGTGC 558
DB 1087 rSerValAlaGlnAlaPheThrArgSerAsnProLeuProAlaArgCysThrGlyHisAl 1107
QY 557 AGCTGTCACGCCGAGCCCTCAGATAGGTACAGCCAGGCGCTTCACATATGTGCTTGT 498
DB 1107 aCysAlaCysProCysProGln----- 1114
QY 497 -GGATCAGGAGCACTTGCTCCATCCAGTCTGAAAAAGCAGCTGCCGA----- 446
DB 1115 -----SerArgProSerCysArgHisValAl 1123
QY 445 -----CCTGGGGGACAGAGTGAAGGTAAAGGCTGTGATGT 408
DB 1123 aglThrGlnSerLeuArgLeuProSerTyrArgGlnAlaCysValGlnGlyValProAl 1143
QY 407 TTCTCTTTCTGGAGGACCAACCAATTCCTTAATTGATGCTGCTTGAAGACAGTATACC 348
DB 1143 aglyValAlaAla-----ThrTrpGlnPr 1151
QY 347 TCTCTCTCAT-----CACCTCAGACCCCTCTGACTTCGATGGGGGATCACTAT 297
DB 1151 OArgGlnHisValCysLeuHisThrHisLeuProPheCysTrpGlyThrValCys 1171
QY 296 CTATCACTCCAGAGAGAGTCTCCGCTCATCTCTGGAAGGGCTTCTCCACATGGCTCAC 237
DB 1171 sArgHisProProProCysSerSerHisSerProTrp----- 1183
QY 236 CTGTCCAGGTGTGTGTTGT-----GACCACAAAGCCCGACACTGGG 192
DB 1184 -----LeuIleGlyThrTrpGlnProProSerHisArgGlyArgThrLeuGl 1199
QY 191 CATG 188
DB 1199 Yleu 1200
  
```

RESULT 2

US-09-467-997-1

Sequence 1, Application US/09467997

Patent No. 6379925

GENERAL INFORMATION:

APPLICANT: KilaJewski, Jan

APPLICANT: Uytendaele, Hendrik

TITLE OF INVENTION: ANGIOGENIC MODULATION BY NOTCH SIGNAL TRANSDUCTION

FILE REFERENCE: 53663-A-PCT-US

CURRENT FILING DATE: 1999-12-20

NUMBER OF SEQ ID NOS: 10

SOFTWARE: Patentin Ver. 2.1

SEQ ID NO. 1

LENGTH: 1964

TYPE: PRT

ORGANISM: mouse

US-09-467-997-1

Alignment Scores:

Pred. No.:	0.00163	Length:	1964
Score:	120.00	Matches:	128
Percent Similarity:	28.72%	Conservative:	38
Best Local Similarity:	22.15%	Mismatches:	193
Query Match:	4.75%	Indels:	220
DB:	4	Gaps:	36

US-09-989-919-15 (1-1397) x US-09-467-997-1 (1-1964)

```

QY 6 TGCACCTGTACCGGAGCGGAGATATCTGCAAGATCCACCG-----CAAGCAGAGTA 59
DB 903 CysIleAspThrGlySerSerTyrPheCysArgCysProProGlyPheGlnGlyLeu 922
QY 60 CCGAGTACAGTGTATCCACAGACAGACAGACCATGCCCGAG----- 97
DB 923 CysGlnAspAsnValAsnProCysGlnProAsnProCysHisHisGlySerThrCysVal 942
  
```

```

QY 98 -----GAAGACTAC-----CGTGCTGGCCATCTTACCAACAGGGAGCTGCTC 142
Db 943 ProGlnProSerGlyValCysGlnCysAlaProGlyValGlnGlnCys--- 961
QY 143 CTTTCAAGTTCACCGCTGAGGCTGATGTCGTGAGAGCCATGCCAGTCCG 202
Db 962 -----SerLysValLeuAspAlaCysGlnSerGln---ProCysHis 974
QY 203 GCGTTGTGTGTCACCAACAGACACCTGAGCAGTGGAGAGAA-----GCC 256
Db 975 -----AsnHisGlyThrCysThrSerArgProGlyGlyPheHisCysAla 989
QY 257 CTTCCAGGAGATGGCAGGACCTCTGTGAGGTTG-----ATAGATACT----- 301
Db 990 CysProProGlyPheValGly-----LeuArgCysGlnGlyAspValAspGluCysLeu 1007
QY 302 GATCCCCCATCGGAAGTCAGAGGGGCTGCT-----GAGTGATGAGAGAGAGTAT 352
Db 1008 AspArgProCysHisProSerGlyThrAlaAlaCysHisSerLeuAlaAsnAlaPheTyr 1027
QY 353 -----ACGTGCTTTCAGGC----- 367
Db 1028 CysGlnCysLeuProGlyHisThrGlyGlnArgCysGlnValGlnMetAspLeuCysGln 1047
QY 368 AGTCAAAATTAGGAGAAATGCTTTCCTCCAGAAAGAAACATCCAGCCCT-----GT 421
Db 1048 SerGlnProCysSerAsnGlyGlySerCys-GluIleThrThrGlyProProProGlyPhe 1067
QY 422 TACCTCTCACTCTGCCCCCAGGTGGCAG----- 452
Db 1067 eInrCysHis---CysProLysGlyPheGlnGlyProThrCysSerHisLysAlaLeuSe 1086
QY 453 -----CTGGTCTTTTTCAGACTGATGAGACCAAGTGTCTTCATCCCAAGACAC 508
Db 1086 rCysGlyLleHisHisCysHisAsnGlyGlyLeuCysLeuProSerPro----- 1102
QY 509 AATATGGAAGCCTCTGGCTGACCT-----ATCTGAGGCTCGCTGACCA-- 554
Db 1103 -----LysProGlySerProProLeuCysAlaCysLeuSerGlyPheGlyGlyProAs 1120
QY 555 -----GCTGACTATCTTCAGACGACTGGGCTTGCTT-----GTGAGGAGAT 555
Db 1120 rCysLeuThrProProAlaProProGlyCysGlyProProSerProCysLeuHisAsnG 1140
QY 596 GACTTGACCTGAGCAGACTGATGTCACCTGGAGAACCCCGAGACAAAGCTAAATCC 655
Db 1140 yThrCysThrGlnThrProGly-----LeuGlyAsnProGlyPheGlnCysThrCysP 1158
QY 656 AGACAGACAGATGTGACAGACAAACGTGCAT----- 689
Db 1158 oProAspSer-----ProGlyProArgCysGlnArgProGlyAlaSerGlyCysGlnG 1176
QY 690 -----AATGCCAAATGTTAAATGTGAGTTTACACGCTTACCT 727
Db 1176 yArgGlyGlyAspGlyThrCysAspAlaGlyCys-----SerGlyProGlyGlyAs 1193
QY 728 ATGGAGC----- 734
Db 1193 pThrAspGlyGlyAspCysSerLeuGlyValProAspProTyrGlyGlyCysProProHis 1213
QY 735 -----TGCTGCTCTTACTCCAGATCATGAG----- 761
Db 1213 sSerGlnCysTyrPheLeuLeuPheArgGlyArgCysHisProGlnCysAspSerGlnG 1233
QY 762 -----GGTATGACTGCTCTTCCAAACCTGTG----- 788
Db 1233 uCysLeuPheAspGlyTyrAspCysGlnLeuProProThrCysIleProAlaTyrAspG 1253
QY 789 -----GGCTGTAAGACACTCAGG 808
Db 1253 nTyrCysArgAspHisPheHisAsnGlyHisCysGlnLysGlyCysAsnAsnAlaGlyC 1273
QY 809 TACTCTCCCACTGGGGGCTGTGCCCCCTCCCTGGAGACGTTCCGTGGGACGCCCATCAC 868

```

```

Db 1273 sGlyTyrAspGlyGlyAspCysArgProGlnGlyLeuAspSerGlnGlyArgProSerLe 1293
QY 869 TGTGTTCAATATGATGTGAGAAATGATAGCTAAAGCCCCCTGCTGCTGCTGCAATGCGACA 928
Db 1293 uAla-----LeuLeuValValLeuArgProPro- 1302
QY 929 GCAGCGGTGGGGGCTGCTGGGGACAAATCCATGCTGAGATGTTCTTCAAGCTTAGGTCT 988
Db 1303 -----AlaLeuAspGlnGlnLeuLeuAlaLeuAlaArgValLeuSerLe 1317
QY 989 GCACAGAGACTTGGCGGGGATGCTCCAGATGTGGTGATTTCTTACCTGGGAGGCT 1048
Db 1317 uThrLeuArgValGly-----LeuThrVal-----ArgLys 1327
QY 1049 ATCTGACCTCCCGACAGGGACACTCCCA----- 1079
Db 1327 sAspSerGlnGlyArgAsnMetValPheProTyrProGlyThrArgAlaLysGlnGluLe 1347
QY 1080 -----GGCCAGCCCGAGGGGTCCAGGGGACAGAGTGCACACCTCAGCATGAGCCA- 1127
Db 1347 uSerGlyAlaArgAspSerSerSerThrProGlnGlnAlaProProThrGlnProLeuG 1367
QY 1128 -----AGACTGGGCTCAGG-----AGCAG 1147
Db 1367 yLysGlnThrGlnGlnSerLeuGlyAlaGlyPheValValMetGlyValAspLeuSerLe 1387
QY 1148 GTGTGTTTGGAGCAGGACTGGGGGCG-----GGGTGGGGCGGGGCTTTCCTTCCTCAT 1204
Db 1387 rCysGlyProGlnHisPheAlaSerArgCysProThrAspSerGlyLeuLeuLeuArgP 1407
QY 1205 TGCTTCAATGAAGCCTCAAGACGCCAAACACAGGCTTTCCTTCCTCCTC 1256
Db 1407 eLeu-----AlaAlaMetAlaAlaValGlyAlaLeuGlnProLeuLeu 1421

RESULT 3
US-08-093-453B-3
; Sequence 3, Application US/08093453B
; Patent No. 5439814
; GENERAL INFORMATION:
; APPLICANT: Frey, Terry K.
; APPLICANT: Dominguez, Geraldina
; APPLICANT: Wang, Chin Yen
; TITLE OF INVENTION: Modified Infectious Rubella Virus
; NUMBER OF SEQUENCES: 13
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Jamie L. Greene, Jones & Askew
; STREET: 191 Peachtree Street, 37th Floor
; CITY: Atlanta
; STATE: Georgia
; COUNTRY: United States
; ZIP: 30303
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50
; COMPUTER: Macintosh
; OPERATING SYSTEM: 7.0
; SOFTWARE: Microsoft Word
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/093,453B
; FILING DATE: 19 JUL 1993
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: U.S. 07/722,334
; FILING DATE: 28 JUN 1991
; ATTORNEY/AGENT INFORMATION:
; NAME: Greene, Jamie L.
; REGISTRATION NUMBER: 32,467
; REFERENCE/DOCKET NUMBER: 07362-0101
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 404 818-3700
; TELEFAX: 404 818-3799
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:

```

LENGTH: 1063 amino acids
 TYPE: amino acid
 STRANDEDNESS: single
 TOPOLOGY: linear
 MOLECULE TYPE: peptide
 HYPOTHETICAL: NO
 ANTI-SENSE: NO
 FRAGMENT TYPE: C-terminal
 ORIGINAL SOURCE:
 ORGANISM: Rubella virus
 STRAIN: Thierien
 US-08-093-453B-3

Alignment Scores:
 Pred. No.: 0.00158 Length: 1063
 Score: 119.00 Matches: 96
 Percent Similarity: 31.65% Conservative: 36
 Best Local Similarity: 23.02% Mismatches: 136
 Query Match: 4.65% Indels: 149
 DB: 1 Gaps: 20

US-09-989-919-15 (1-1397) x US-08-093-453B-3 (1-1063)

QY 1184 GCCCAACCCCGCCGAGCTCTGCGTCAACC-----ACACCTGCTCCCTGACCC 1134
 DB 77 AlAProProProProGluGluArgGlnGlnThrArgSerGlnThrProAlaProLysPro 96
 QY 1133 CAGCTTGCTCACTGAGAGTGTCACCTGCTCCCTGACCCCTGGGCTGGCTGGAG 1074
 DB 97 SerArg-----AlaPro-----ProGlnGlnProGln----- 105
 QY 1073 TGTCCTCTGCGAGGTCAAGATAGCTCCCGAGTACAG-----AAT 1029
 DB 106 -----ProProArgMetGlnThrGlyArgGlyGlySer 116
 QY 1028 CACCCACATCTGAGAGATCCCGCCCAAGTCTCTGCTCCAGCTTACAGTGAAGA-- 972
 DB 117 AlAProArgProGluLeuGlyProProThrAsnProGlnAlaAlaValAlaArgGly 136
 QY 971 -----ACACTCCAGATGATGTGCTCCAGCAGCCCGCCCGCTGCTG-- 926
 DB 137 LeuArgProProLeuHisAspPro--AspThrGln-AlAProThrGlnAlaCysValThr 155
 QY 925 -----GGCATGTGACGACGACGACGAGGGCTTACCTGATCTCACTATTGAAC 870
 DB 155 rSerTrpLeuTrpSerGlnGlyGlnGlyAlaValPheTyrArgValAspLeuHisPhe 175
 QY 869 AGTATGGGGGTGCTCCAGCGAACCCTCCAGCGA----- 836
 DB 175 rAsnLeuGlyThrProProLeuAspGluAspGlyArgTrpAspProAlaLeuMetTyrAs 195
 QY 835 -----GGGCGACAGCCCGCCAGTGGGAGACTAGCTGAGCTTGCTTACAGCCACAG 783
 DB 195 nProCysGlyProGluProPro--AlaHisValAlaArgAlaTyrAsnGlnProAlaG 214
 QY 782 GTTGAAGAGCAGATACCCCATGATCTCTGAGCTAGAGCCAGAGATCCCATACTA 723
 DB 214 yAspValArgGlyValTrpGlyLysGlyGluArgThrTyrAlaGlnGlnAspPheArgVa 234
 QY 722 GCGGTGTAACCTGACATTTAACATTGGCATTTGTCAGCTTGCTGCTGCTGACATCTG 663
 DB 234 lGlyGly-----ThrArgTrpHisArgLeu-----ArgMetP 246
 QY 662 TCTGTCTGGATGTAGCTTTGTCTGACAGGGTTCCAGGTGACATGACGTGCTGCCAGT 603
 DB 246 oValArgGlyLeuAspGlyAspSerAlaProLeuProPro--HisThrArgGluArgG 265
 QY 602 GCAAGTCACTCCCTCCACAGGACAGCCAGCTGCTGAGATAGTACGTGTACGGCG-- 545
 DB 265 eGluThrArg-----SerAlaArgHisProTyr 274
 QY 544 -AGCCCTAGATAGTACGACAGAGGCTTCACATATGTGCTTGTGGATCAGCAGAC 486
 DB 544 ----- 486

DB 274 pArgLLeuRhpheGlyAlaProGlnAlaPheLeuAlaGlyLeuLeuLeuAlaThrValAl 294
 QY 485 CACTTGGCTCCATCCAGTCTTTGAAAAAGACCAAGCTCCGAGCTGGGGGCGAGAGTGGA 426
 DB 294 a-ValGlyThrAlaArgAlaGlyLeuGlnProAlaArgAlaAspMet----- 308
 QY 425 GGTAAACAGGCTGATGTTCTCTTTCTTGAGAGGCAAGCANTCTCCTAATTGACTGC 366
 DB 308 ----- 308
 QY 365 CTTGAAAGACAGTATACCTCTCTCATACCTACAGACCCCTCTGACTTCCAGTGGAG 306
 DB 309 -----AlaAlaProProThr-LeuProGlnPro 317
 QY 305 GATCACTATCTATCAACTCCAGAGAGGTCTGCACTCTCCCTTGAAGGGCTTCTCCA 246
 DB 318 -----Pro 318
 QY 245 CTGGCTCACCTGTCCAGGTGCTGTGTGATACCAAGCCCGACACTGGGCGATGCG 186
 DB 319 CysAlaHisGlyGlnHisTyrGlyHisHisHisGlnLeuProPheLeuGlyHisAsp 338
 QY 185 TCTCAAGACATCCACAGCTTCAGCCAGGTGAACATGAAGAGCAGCTCCGTGCT 126
 DB 339 GlyHisHis-----GlyGlyThrLeuArgVal 347
 QY 125 GGTAGATGGCCAGCAGCGTAGT-----CTTCCCTGGGGGATGG 87
 DB 348 GlyGlnHisTyrArgAsnAlaSerAspValLeuProGlnHisStrpLeuGlnGlyTyr 367
 QY 86 TGCTGTCTGGGA-----TACACTGTGTACTGGTACTGC 54
 DB 368 GlyCysTyrAsnLeuSerAspTrpHisGlnGlyThrHisValCys 382

RESULT 4
 US-07-960-389-2
 ; Sequence 2, Application US/07960389
 ; Patent No. 5705611
 ; GENERAL INFORMATION:
 ; APPLICANT: HAYASHIDA, Kasuhiro;
 ; TITLE OF INVENTION: Human GM-CSF Receptor Component
 ; NUMBER OF SEQUENCES: 2
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Schering-Plough Corporation
 ; STREET: 2000 Galloping Hill Road
 ; CITY: Kenilworth
 ; STATE: New Jersey
 ; COUNTRY: USA
 ; ZIP: 07033
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy Disc
 ; COMPUTER: Apple Macintosh
 ; OPERATING SYSTEM: System Software 7.1
 ; SOFTWARE: Microsoft Word 5.1a
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/07/960,389
 ; FILING DATE: 07-JAN-1993
 ; CLASSIFICATION: 800
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: 554,745
 ; FILING DATE: 18-JUL-1990
 ; APPLICATION NUMBER: PCT/US 91/04846
 ; FILING DATE: 16-JUL-1991
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Blasdale, John H. C.
 ; REGISTRATION NUMBER: 31,895
 ; REFERENCE/DOCKET NUMBER: DX01430
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: (908) 298-2902
 ; TELEFAX: (908) 298-5388
 ; INFORMATION FOR SEQ ID NO: 2:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 897 amino acids

TYPE: amino acid
 STRANDEDNESS: single
 TOPOLOGY: linear
 MOLECULE TYPE: protein
 FEATURE:
 OTHER INFORMATION: Human GM-CSF receptor; Signal Sequence: -17 to -1; Transmem
 US-07-960-389-2

Alignment Scores:
 Pred. No.: 0.00185 length: 897
 Score: 118.00 Matches: 104
 Percent Similarity: 34.37% Conservative: 51
 Best Local Similarity: 23.06% Mismatches: 152
 Query Match: 4.61% Indels: 144
 Gaps: 22

US-09-989-919-15 (1-1397) x US-07-960-389-2 (1-897)

QY 1221 GGGTTTCATTGAAAGCAATGAGGAGAAAGCCCGCCGCCACCCCGCCCGAGTCTCT 1162
 DB 464 GYIleTyrclyrArgleuAArglyrTprgluGluylleProAenProserLys 483
 QY 1161 GGC-----TCAAAACCAACCTGCTCCCTACCCCACTCTTGGCTATGCTAGAGG 1111
 DB 484 SerHisLeuPheGlnasnGlySerAlaGluLeuTprProglYserMetSerAlaPhe 503
 QY 1110 TGCACCTTCGCGCT-----GACCCCTGGGCTGGCGCTG-----GGG 1075
 DB 504 ThrSerGlySerProHisiGlnGlyProTprGlySerArgPheProGluLeuGly 523
 QY 1074 GTGTCCCTGCTC-----GGGAGGTGAGAGATGCTCCCGCAGTACAGATCACCACCA 1021
 DB 524 ValPheProValGlyPheGlyAspSerGluVal-----SerProLeu 537
 QY 1020 TCTTGAGAGATCCCGCCCAAGTCTCTCTGTCAGACTTAAGCTGAGAGAACTCCACGA 961
 DB 538 ThrIleGluAspProHisiValCyAspProProSerGlyProAspThrThrProAla 557
 QY 960 TGGATT---GTCCCAAG-----CAGCCCGCCAGCGCTGCTGGGCA 922
 DB 558 AlaSerAspLeuProThrGluGlnProProSerProGlnProGluProAlaAlaSer 577
 QY 921 TGTGACGACGACGACGAGGGCTTTAGCTA----- 891
 DB 578 HisThr-ProGluLysGlnAlaSerSerPheAspPheAsnGlyProTyrLeuLysProPr 597
 QY 890 -CATTTCTACACTATTGACACAGTATGGGCTGCCACGGAACCGTCCCGAGGAGGG 832
 DB 597 ohIsSerArgSer-LeuProAspIleLeuGlyLnpProGlnGlnGlyGlyS 617
 QY 831 CACAG-----CCCCAGTGGGAGACTACCTGAGCTTGTACAGCCACAGGGGTGG 778
 DB 617 erGlnLysSerProProGlySerLeuGlyLnuTyrLeuCyAsu---ProAlaGlyGlyG 636
 QY 777 AGAGGAGATACACCCCATGATCTCTGACTAGAGACCGACGATCCCATAGCTAGGCTG 718
 DB 636 lnaValGlnLeuValPro----- 641
 QY 717 GTAACCTACATTTAACATTGTCATTATTCACGTTGTCTGTGTACATCTGTCTGT 658
 DB 641 ----- 641
 QY 657 CTGGGATGTTAGTTTGTCTGACGGGTTCCAGGTGACATGACGTGCTCCAGTCCAG 598
 DB 642 -----LeuAlaGlnAlaMetGlyProGlyLys-----AlaValGluValGlu 656
 QY 597 TCACCTCCCTCAGACGAAAGCCAGCTGTGAGAGATGACAGTGTGTCGACGACCTC 538
 DB 656 rArgArgProSerGlnGly-----AlaAlaGlySerProSerLeuG 669
 QY 537 AGATAGGTGACGACGAGGCTTTCACATATGTGTCTTTGGATCAGGAGCACCTTGGC 478
 DB 669 luserGlyGly-----GlyProAlaProProAlaL 679

QY 477 TCATTCAGCTTTGAAAAAGACCAGCTGCCGACTGGGGGAGAGGTGAGAGTACAG 418
 DB 679 euGlyPro-----ArgValGlyGlyGlnAspGlnLysAspSerP 692
 QY 417 GCGTGAATTTTCTCTTTCTTGAGGACGACCACTTCTCCCTAATTGACTGCTTGAAGA 358
 DB 692 roValAlaIlePrometSerSerGlyAspThrGluAspProGlyValAlaSerGlyTyTy 712
 QY 357 CACGATACCTCTCTCTATCACTCAGACACCCCTCTACCTTCCATGGGAGATCACTA 298
 DB 712 alSer-SerAlaAspLeuVal-----PheThrProAsnSerGlyAlaSer 726
 QY 297 TCTATCACTC-----CCAGAGGTCTCCACTCTCTCT 262
 DB 727 SerValSerLeuValProSerLeuGlyLeuProSerAspGlnThrProSerLeuCySp 746
 QY 261 GGAAGGCTTCT---CCACTGCTCCTCCTGTCACGTG----- 226
 DB 747 GlyLeuAlaSerGlyProProGlyAlaProGlyProValLysSerGlyPheGluGlyTy 766
 QY 225 GTCTGCTGTGTACACCAAGGCCGACACTGGGATGGCTCTCACAGACATCCACAGC 166
 DB 767 ValGluLeuProProIleGluGlyArg-SerProArgSerProArgAsnAsnProValP 786
 QY 165 T-----CAGCCAGTTGAACTGAAGAGGACGACTCCCG 130
 DB 786 oProGluAlaLysSerProValLeuAsnProGly-----GluTr 799
 QY 129 TGGTGTAGATGCGCAGCAGCGAGTACTTCTTGGGGAGATGCTGTCTGAGATAC 70
 DB 799 gProAlaAspValSerProThrSerProGlnProGluGlyLeuLeuValLeuGlnGln 819
 QY 69 TGTACTCGTACTGCTGCTTGGC 45
 DB 819 lGlyAsp---TyrCySphLeuPro 826

RESULT 5
 US-09-041-886-23
 Sequence 23, Application US/09041886
 Patent No. 6235872
 GENERAL INFORMATION:
 APPLICANT: Bredesen, Dale E.
 TITLE OF INVENTION: Prosoplopic Peptides, Dependence
 NUMBER OF SEQUENCES: 72
 CORRESPONDENCE ADDRESS:
 ADDRESSER: Campbell & Flores LLP
 STREET: 4370 La Jolla Village Drive, Suite 700
 CITY: San Diego
 STATE: California
 COUNTRY: United States
 ZIP: 92122
 COMPUTER READABLE FORM:
 MEDIUM TYPE: floppy disk
 OPERATING SYSTEM: IBM PC compatible
 SOFTWARE: Patent in Release #1.0, Version #1.25
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/09/041,886
 FILING DATE:
 CLASSIFICATION:
 ATTORNEY/AGENT INFORMATION:
 NAME: Campbell, Cathryn A.
 REGISTRATION NUMBER: 31,815
 REFERENCE/DOCKET NUMBER: P-LJ 2626
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (619) 535-9001
 TELEFAX: (619) 535-8949
 INFORMATION FOR SEQ ID NO: 23:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 1185 amino acids

TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-09-041-886-23

Alignment Scores:

Pred. No.:	Length:	Matches:	Conservative:	Mismatches:	Indels:	Gaps:
Score: 117.00	1185	92	34	140	125	20
Percent Similarity: 32.23%						
Best Local Similarity: 23.53%						
Query Match: 4.57%						

US-09-989-919-15 (1-1397) x US-09-041-886-23 (1-1185)

```

QY 1190 GCGCCGCGCCGACCC----- 1176
DB 366 AlaProAlaProBrometArpPheProTySerSerSerSerSerSerAlaAlaAla 385
QY 1175 -----CCGCCCAAGTCTGCTCAACACCA 1149
DB 386 SerSerSerSerSerSerSerSerSerSerAlaSerProPheProAlaSerGlnAlaLeu 405
QY 1148 CCGCT-----CCCTGACCCGAGCTTGCTCAGTGAAGTGACCT 1104
DB 406 ProSerTyProHisSerPheProProProThSerLeuSerValSerGlnProPro 425
QY 1103 CTGCGCCCTGACCCCTGGGCTGCTGGAGTGTCCTGTCGGAGAGGTCAAGATAGCT 1044
DB 426 LysTyThrGlnProSerLeuProSerGlnAlaValTrpSerGlnGlyProProPro 445
QY 1043 CCCCAG-----GTACAGATCACCCACATCTTGAGAGATCCCGCCAG 999
DB 446 ProProTyGlyAlaGluLeuAlaAsnSerAsnAlaHisProGlyProPheProPro 464
QY 998 TCTCTCTCCAGACTTAACTAGAGAGAACTCCAGCATGATGTCGCCACGACGACCC 939
DB 465 -----SerThrGlyAlaGlnSerThrAlaHisProProValSerThrHisHis 481
QY 938 CACCGCTGCTGTGTGATGTGACAGACAGACAGAGGCTTAACTACAT----- 888
DB 482 HisHisGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 501
QY 887 -----TTCACACTTTTGAACACAGTGAAGGCTGCC 855
DB 502 SerGlyProProProGlyAlaPheProHisProLeuGlu-----GlyGlySerSer 519
QY 854 CAGGGAACCGTCCAGAGGAGGAGGACAGACGCGGAGACTAGCTGAGCTTCTT 795
DB 520 HisHisAlaHisPro--TyrAlaMetSerPro-----SerLeuGlySerLeu 534
QY 794 ACACCCCAAGAGGTTGAGAGGAGCATATACCCCATGATCTTGAGACTAGAGCCAGCA 735
DB 535 ArgProTyRProProGlyProAlaHisLeuProPro----- 546
QY 734 GTCCCATAGCTAGGCTGTAACTCACATTTTAACATTGGCATTATTCAGCTTGCC 675
DB 547 -----ProHis-----SerGlnValSer 552
QY 674 TGGTCACATCTGTCTGTCTGGATGTAGCTTTGTCGAGGAGGTCCAGAGTACATGC 615
DB 553 TySerGlnAlaGlyPro-----AanglyProProValSerSer 565
QY 614 AGTGTGACAGTGAAGTCACTCCCTCCAGAGGAGCCAGCTGCTGAGATAGTACGC 555
DB 566 SerSerHisSerSerSerSerSerThnSerGlnGlySerTyRProCys----- 580
QY 554 TGGTCAGCCGAGCTCTGAG-----ATAGTTCAGCCAGAGGCTTTCACATATGCT 504
DB 581 SerHisProSerProSerGlnGlyProGlnGlyAlaProTyRProPheProProValPr 600
QY 503 CTGTGTGATCAGGAGGACCACTGGCTCCATCCATTCCTTGAAGAACCACTGCCGAC 444

```

```

DB 600 cThrValThrThrSerSerSerAlaThrLeuSerThrValIleAlaThrValAlaSerSerPr 620
QY 443 TGGGGGCGAGAGGTGACAGAGTAACAGGGCTGAGTCTTCTCTGGA----- 395
DB 620 AlaGlyTyRlys-----ThrAlaSerProProGlyProProProTy 634
QY 394 -GGCAAGACATTCCTCTAATTGACTGCTTGAAGACAGATATACCTCTCTCATCA 336
DB 634 rGlyLysArg--AlaProSerProGlyAlaTyRlysThrAlaThrProProGlyTyRly 653
QY 335 CCTCAGACACCCCTCTGACTTCGATGGGAGATCATATCATCACTCCAGAGAGGTC 276
DB 653 sProGlySerProProSerPheArgThrGly-----ThrProPro----- 666
QY 275 CTGCATCTCTCTTGAAGGCTTCTCCACTGCTCACTGCTCATCTGCTGAGTGTGTTGG 216
DB 667 -----GlyTyRArgGly-----ThrSerProProAlaGlyProGly 678
QY 215 TGACCACAAA-----GGCCGACACAGTGGGC 191
DB 678 YThrPheLysProGlySerProThrValGly 688

```

RESULT 6

US-09-046-158A-2
Sequence 2, Application US/09046158A
Patent No. 6187552

GENERAL INFORMATION:

APPLICANT: Roberts, Steven L.
APPLICANT: Kayles, Paul S.
TITLE OF INVENTION: METHOD FOR IDENTIFYING INHIBITORS OF
TITLE OF INVENTION: JAK2/CYTOKINE RECEPTOR BINDING
NUMBER OF SEQUENCES: 30
CORRESPONDENCE ADDRESS:

ADDRESSEE: Pharmacia & Upjohn Co., Intellectual Property
ADDRESSEE: Legal Services
STREET: 301 Henrietta Street
CITY: Kalamazoo
STATE: MI
COUNTRY: USA
ZIP: 49001

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/046,158A

FILING DATE:

CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:

NAME: Darinley Jr., James D.
REGISTRATION NUMBER: 33,673
TELECOMMUNICATION INFORMATION:
TELEPHONE: 616/833-2210
TELEFAX: 616/833-8897

TELEX: 224401

INFORMATION FOR SEO ID NO: 2:

SEQUENCE CHARACTERISTICS:
LENGTH: 433 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide

US-09-046-158A-2

Alignment Scores:

Pred. No.:	Length:	Matches:	Conservative:	Mismatches:	Indels:	Gaps:
Score: 116.00	433	103	50	149	144	22
Percent Similarity: 34.30%						
Best Local Similarity: 23.09%						
Query Match: 4.53%						

US-09-989-919-15 (1-1397) x US-09-046-158A-2 (1-433)

QY 1206 CAATGAGGAGAAAGGCCCCGCCCCAGAGCTCTGCG-----TCA 1156
 Db 5 ArgLeuAkrgrlystrprgluglylAlaProAsnProSerlysserthlaleuPheGln 24
 QY 1155 AACCAACCTGCTCCCTGAGACCCAGCTTGGCTCAGTACAGGTGACCTGCCCCCT 1096
 Db 25 AsnGlySerAlaIleuTrpProProGlySerMetSerAlaPheThrSerGlySerPro 44
 QY 1095 -----GACCCCTGGGCTGGCTG-----GAGAGTCCCTGTC--- 1063
 Db 45 ProHisGlnIleProTrpGlySerArgPheProGlyLeuGlnGlyAlaPheProValGly 64
 QY 1062 ---GGAGAGTCAGAGATAGCTCCCGAGTACAGAAATCACACCAATCTGAGACATCCCC 1006
 Db 65 PheGlyAspSerGlyVal-----SerProLeuThrIleGlyLeuAspPro 78
 QY 1005 CGCCAAAGTCTCTGCTGAGCCTTAAGCTGAGAGAAACATCCACGATGATT---GTCCCC 949
 Db 79 LysHisValCysAspProProSerGlyProAspThrThrProAlaAlaSerAspLeuPro 98
 QY 948 AGC-----CAGCCCCACCGCTGCTGTCGATGTGACAGACAGCA 907
 Db 99 ThrGlnGlnProProSerProGlnProGlyProProAlaAlaSerHisThr-ProGly 118
 QY 906 GCAGGGGCTTAACTA-----CATTCCTCACACTAT 877
 Db 118 sgnAlaSerSerPheAspPheAsnGlyProTyrLeuGlyProProHisSerArgSer-L 138
 QY 876 TGAACACAGTATGGGGCTGCCCCAGAACCTGCCAGGAGGGGAGCAG-----CCCC 823
 Db 138 eurProAspIleLeuGlyGlnProGlnProProGlnGlnGlyGlySerGlnLysSerPro 158
 QY 822 CAGTGGGGAGACAGCTGAGCTTGTCTTACAGCCCAAGGCTTGAGAGAGCACTCATAC 783
 Db 158 robProGlySerLeuGlyTyrLeuCysLeu---ProAlaGlyGlyGlnValGlnLeuValP 177
 QY 762 CCCATGATTCCTGAGTACAGGAGCAGAGTCCATAGCTAGCTGTAACATCATATTT 703
 Db 177 ro----- 177
 QY 702 AACATTGGCATATTGACAGCTTGTCTGTCATCATCTGTCTGTGGAGTTAGCTT 643
 Db 178 -----LeuAlaG 180
 QY 642 TGTCTGAGGGGTTCCTCAGGTGACATGACATGCTGCCAGTGCATCTCCTCCACAG 583
 Db 180 lnaIaMetGlyProGlyGln-----AlaValGlyValGlyArgArgProSerGlnG 197
 QY 582 GCAAGCCACGCTGCTGAGTACAGCTGTCAGCCGAGCCCTCAGATAGTGCAGCCAG 523
 Db 197 ly-----AlaAlaIleSerProSerLeuGlnSerGlyGly--- 208
 QY 522 AGGCTTCACATATGTGCTTGTGTTGGTTCAGGAGCAACCTTGCTTCATCCATCTTGA 463
 Db 209 -----GlyProAlaProProAlaLeuGlyPro----- 217
 QY 462 AAAAGACAGCTCCGACCTGGGGGAGAGGTGAGAGTAAACAGGGCTGAGTGTTC 403
 Db 218 -----ArgValGlyGlyGlnAspGlnLysAspSerProValAlaIleProM 223
 QY 402 TTTCTGAGGAGCAAGACCATTTCTCCCTAATTGACTGCTTGAAGACACGATACCTCTCT 343
 Db 233 etSerSerGlyAspThrGlnAspProGlyValAlaSerGlyTyrValSer-SerAlaAsp 252
 QY 342 CTGATCACTCAGACACCCCTGACTCCGATGGGGGAGTCACTATCTATCACTC--- 286
 Db 253 LeuVal-----PheThrProAsnSerGlyAlaSerSerValSerLeuVal 267
 QY 285 -----CAGAGAGTCTGCTGACATCTTCCTTGAAGAGGCTTCT--- 250
 Db 268 ProSerLeuGlyLeuProSerAspGlnThrProSerLeuGlyLeuAlaSerGly 287

QY 249 CCCACTGGCTACCTGTCCAGTGT-----GTCTGTTGTGAC 211
 Db 288 ProProGlyAlaAlaProGlyProValLysSerGlyPheGlnGlyTyrValGlnLeuProPro 307
 QY 210 ACAAGAGCCGACACTGGGATGCTCTACAGACATCCACAGCT--- 165
 Db 308 lIeGlnGlyArg-SerProArgSerProArgAsnProValProProGlnAlaLysSe 327
 QY 164 -----CAGCCAGTTGAACACTGAAGAGGAGCAGCTCCGTGTGTGATGTC 115
 Db 327 rProValLeuAsnProGly-----GluArgProAlaAspValSe 340
 QY 114 CAGACCGGTAGTCTTCTGGGGATGTGCTGTCTGAGTACACTGTAAGTACTG 55
 Db 340 rProThrSerProGlnProGlnGlyLeuLeuValLeuGlnGlnValGlyAsp---TyrCy 359
 QY 54 CTGCTTGGCG 45
 Db 359 spHeuLeuPro 362

RESULT 7

US-08-783-774-2

; Sequence 2, Application US/08783774
 ; Patent No. 6054130

GENERAL INFORMATION:

APPLICANT: Spaete, Richard
 APPLICANT: Jackman, Winthrop

TITLE OF INVENTION: NON-SPLICING VARIANTS OF
 ; TITLE OF INVENTION: GP350/220

NUMBER OF SEQUENCES: 19
 CORRESPONDENCE ADDRESS:

ADDRESSER: Pennie & Edmonds
 STREET: 1155 Avenue of the Americas

CITY: New York
 STATE: NY

COUNTRY: USA
 ZIP: 10036/2711

COMPUTER READABLE FORM:
 MEDIUM TYPE: Diskette

COMPUTER: IBM compatible
 OPERATING SYSTEM: DOS

SOFTWARE: FastSeq Version 2.0
 CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/783,774
 FILING DATE: 15-JAN-1997

CLASSIFICATION: 435
 ATTORNEY/AGENT INFORMATION:

NAME: Coruzzi, Laura A.
 REGISTRATION NUMBER: 30,742

REFERENCE/DOCKET NUMBER: 7682-037
 TELECOMMUNICATION INFORMATION:

TELEPHONE: 212-790-9090
 TELEFAX: 212-869-8864

TELEX: 66141 PENNIE
 INFORMATION FOR SEQ ID NO: 2:

SEQUENCE CHARACTERISTICS:
 LENGTH: 907 amino acids

TYPE: amino acid
 STRANDEDNESS:

TOPOLOGY: unknown
 MOLECULE TYPE: protein

US-08-783-774-2

Alignment Scores:

Pred. No.:	0.00378	Length:	907
Score:	115.00	Matches:	90
Percent Similarity:	32.77%	Conservative:	46
Best Local Similarity:	21.69%	Mismatches:	117
Query Match:	4.49%	Indels:	162
DB:	3	Gaps:	18

US-09-989-919-15 (1-1397) x US-08-783-774-2 (1-907)

```

QY 1182 CCCACCCGCCCCAGT---CCTGGCTCAACACAGC----- 1147
DB 442 ProAsnThrThrThrglyLeuProSerSerThrhialaProThrsanleuthralaPro 461
QY 1146 TGTCCCTGACCCAGCTTGCTGAGTGTGACCTGTGACCTGCTGACCCCTG 1087
DB 462 AlaserThrglyProThrhialaSerThrhialaSerPro----- 476
QY 1086 GCTGGCTGGAGTGTGCTGCGGAGGTGACAGATAGCTCCCGAGGTACAGATCA 1027
DB 477 -----ThrProalaglyThrThrseryalaserProvalThrProser 491
QY 1026 CCCACATCTGAGCATCCCCGCCAGATCTCTGTCCAGACCTTAAGCTGAGAGAACCT 967
DB 492 ProSerProThrsapasnnglyThrglySerlyalProaspmetThrsSerThrs 511
QY 966 CCAAGATGATTTGCCAGCAGCCCGCTGCTGGCATGTGACAGCAGAGCA 907
DB 512 Pro-----ValThrThrProThrProasmlaThrsSerProThrProalaval 527
QY 906 GCAGGGCTTTAGTACATTCACACTATTGAACAGATGAGGGCTGCCACGAGAC 847
DB 528 Thr----- 528
QY 846 GCTCCAGGAGGGGACAGCCCGCTGAGGAGACTAGCCTGAGCTTGTTCAGCCCA 787
DB 529 -----ThrPro 530
QY 786 CAGGTTGAGAGGAGCATACCCCATGATT----- 754
DB 531 ThrProasmlaThrsSerProThrProalavalThrThrProThrProasmlaThrs 550
QY 753 CTTGAGTAGAGCAGCAGCTCCAGTCTGAGGTGTAACATTTTAACATTGG 694
DB 551 ProThreuglyLysThrsSerProThrservalThrThrProThrProasmla----- 568
QY 693 CATATTGACAGTTTGCTGCTGTCACATCTGTCTGTGGAGTGTGCTTGTCTGAC 634
DB 569 -----AlaThrsSerProThreuglyLysThr-----SerPr 579
QY 633 GGGTTCCAGGTAGCATGAGTGTGCGCAGTGCAGTCACTCCCTCCACAGAGCAG----- 578
DB 579 oThrservalThrThrProThrProasmlaThrsSerProThreuglyLysThrs 599
QY 577 -----CCAGCTGCTGAGGATGAGTCTGCTGAC 547
DB 599 erProThrservalThrThrProThrProasmlaThrglyProThrValglyLut 619
QY 546 CGAGCCTCAG----- 536
DB 619 hrSerProGlnAlaAsnAlaThrsanhlThrhleuglyLysThrsProThrProval 639
QY 535 --ATAAGTCCAGCAGAGCTTTCACATATGTGCTTGTGGATCAGAGCAGCTGGC 478
DB 639 alThrsGlnProLysAsnAlaThrserval-----Thrhrgly 654
QY 477 TCCATTCAGCTTAAAAAGAC-----AGTCCGACCTGGGGGCGAGAGGT 430
DB 654 lnhlsanlleThrsSerSerThrsSerSerMetSerLeuArgProSer----- 670
QY 429 GAGAGGTAAAGGGCTGATTTCTCTTTGAGAGCAGACCATCTCCCTAATTGA 370
DB 671 -----SeranProGlnThrLysSerProSerThrs 681
QY 369 CTGCTTGAAGACAGTATACCTCTCTCATCACTCAGACCCCTGACTTCGAT 310
DB 681 erapasnSerThrsThrsHlmeCProleuthrservalahlspro-----Thr 697
QY 309 GGGGATCACTATTCATCACTCCAGAGAGTCTGACATCTCCCTGAGAGGCTTCT 250
DB 698 GlyglyLys-----Asnle-ThrglnValThrProalaserlle----- 710

```

```

QY 249 CCCACTGCTACCTGTCCAGGTGTGTGTGTGACCAAGAGCCGACACTGGCA 190
DB 711 -----SerThrhshi 714
QY 189 TGGCTTCACAGACATCCACAGCTTCAGCCAGGTGAACTGAAAGAGGAGCTCCG 130
DB 714 svalSerThrsSerProGlnProArgProGlyThrThr-----SerGlnAlase 731
QY 129 TGTGTGTAGATGCGCCACAGGCTAGTCTTCCTGG 93
DB 731 rGlyProGlnAsnSerSerThrsThrhlyserProGly 743

```

RESULT 8

US-09-328-599A-1

Sequence 1, Application US/09328599A

Patent No. 6432679

GENERAL INFORMATION:

APPLICANT: MOND, James J. and Lees, Andrew

TITLE OF INVENTION: Enhancement of B Cell Activation by

TITLE OF INVENTION: Co-ligation of Receptors for Antigen and Complement C3d

TITLE OF INVENTION: Using EBV gp350/220 or EBV gp350/220 Peptide Adjuvants

NUMBER OF SEQUENCES: 2

CORRESPONDENCE ADDRESSES:

ADDRESSEE: Finnegan, Henderson, Farabow, Garrett &

ADDRESSEE: Dunner, L.L.P.

STREET: 1300 I Street, N.W., Suite 700

CITY: Washington

STATE: DC

COUNTRY: USA

ZIP: 20005

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patent in Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/328,599A

FILING DATE:

CLASSIFICATION:

ATTORNEY/AGENT INFORMATION:

NAME: Fordis, Jean B.

REGISTRATION NUMBER: 32,984

REFERENCE/DOCKET NUMBER: 04995, 6025-00000

TELECOMMUNICATION INFORMATION:

TELEPHONE: (202)408-4000

TELEFAX: (202)408-4400

INFORMATION FOR SEQ ID NO: 1:

SEQUENCE CHARACTERISTICS:

LENGTH: 907 amino acids

TYPE: amino acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULAR TYPE: protein

US-09-328-599A-1

Alignment Scores:

Pred. No.: 0.00378

Score: 115.00

Percent Similarity: 32.77%

Best Local Similarity: 21.69%

Query Match: 4.49%

DB: 4

Gaps: 18

US-09-989-919-15 (1-1397) x US-09-328-599A-1 (1-907)

```

QY 1182 CCCACCCGCCCCAGT---CCTGGCTCAACACAGC----- 1147
DB 442 ProAsnThrThrThrglyLeuProSerSerThrhialaProThrsanleuthralaPro 461
QY 1146 TGTCCCTGACCCAGCTTGCTGAGTGTGACCTGTGACCTGCTGACCCCTG 1087
DB 462 AlaserThrglyProThrhialaSerThrhialaSerPro-----ThrsPro----- 476

```

OY	1086	GCTGGCTGGAGAGTGTCCCTGTGGAGAGGTGAGATGAGATAGTACCTCCCAAGGTGACATACAA	1022
Db	477	-----ThrProAlaGlyThrThrSerGlyAlaSerProValThrProSer	491
OY	1026	CCCAATCTGGAGAGCATCCCGCCCAAGATCTCTCTGTCCAGACCTTAAGCTGAGAGAACT	967
Db	492	ProSerProThrPaspAsnGlyThrIuSerLysAlaProAbpMetThrSerSerThrSer	511
OY	966	CCAGCATGATGATTTGCCACGACGAGCCCCACCGCTGTGGGTGGCATGTGCACAGACAGCA	907
Db	512	Pro-----ValThrThrProThrProAsnAlaThrSerProThrProAlaVal	527
OY	906	GAAGGGGCTTTAGTACATTTCTCAACATTTGAACAAGTATGGGGCTGCCACAGAAC	847
Db	528	Thr-----	528
OY	846	CGTCCCAAGAGAGGAGCAGACCCCGAGTGGGAGACTAGACCTGAGCTTTACAGCCCA	787
Db	529	-----ThrPro	530
OY	786	CAGGTTGAGAGGAGCATATCCCATGATT	754
Db	531	ThrProAsnAlaThrSerProThrProAlaValThrThrProThrProAsnAlaThrSer	550
OY	753	CTGGACACTAGAGGACGAGTCCCATAGTACGTAGCGTGAATCATCATTTTAACATTGG	694
Db	551	ProThrLeuGlyLysThrSerProThrSerAlaValThrThrProThrProAsn-----	568
OY	693	CATTATTCACAGCTTTGTCTGTGCATCTGTCTGTGGATGTGTACCTTTGTCTGCAG	634
Db	569	-----AlaThrSerProThrLeuGly-LysThr-----SerPro	579
OY	633	GGGTTCCAGGTGACATGACATGCTGTCCAGATGCAAGTACATCTCTCCACAGGCAAG---	578
Db	579	oThrSerAlaValThrThrProThrProAsnAlaThr-SerProThrLeuGlyLysThrS	599
OY	577	-----CCAGCTGTGAGAGATGATGACGTGTGCAC	547
Db	599	erProThrSerAlaValThrThrProThrProAsnAlaThrGlyProThrAlaGlyLut	619
OY	546	CGAGGCTTCAG-----	536
Db	619	hrSerProGlnAlaAsnAlaThrPaspAsnGlyThrLeuGlyLysThrSerProThrProVal	639
OY	535	--ATAGTTCAGGACAGAGGCTTCAATATGTGTCTTGTGGATCAGAGGACACTTGGC	478
Db	639	alThrSerGlnProLysAsnAlaThrThrSerAlaVal-----ThrThrGlyG	654
OY	477	TCCATCCAGTCTTGA AAAAGACC-----AGCTGCCAGCTGGGGGCGACAGGT	430
Db	654	lnhAsnAlaLehThrSerSerSerThrThrSerMetSerLeuThrProSer-----	670
OY	429	GAGAGGTAAACGGGCTGATGTTTCTTTTTCGAGGGAAGACAACTTGTCCCTAATTGA	370
Db	671	-----SerAsnProGlnThrLeuSerProSerThrS	681
OY	369	CTGACCTTGAAGAAGATACCTCTCTGTCAATCACTAGACACCCCTTGACTTCCGAT	310
Db	681	erAspAsnSerThrSer-HisMetProLeuLeuThrSerAlaHisPro-----Thr	697
OY	309	GGGGATCACTATCTATATCAACTCCACAGAGGTCTGCCATGTCCCTTTGAAGGGCTTCT	250
Db	698	GlyGlyGlu-----AsnIle-ThrGlnAlaThrProAlaSerIle-----	710
OY	249	CCCACTGGCTACCTGTCCAGGTGGTGTGGTGGTGGACACAAAGGCGCGAAGTGGGCA	190
Db	711	-----SerThrHisHis	714
OY	189	TGGCTCTACAGATCCACAGACCTTCAAGGAGTTGAACATGAAAGAGGACGACTCCG	130
Db	714	svAlSerThrThrSerSerProGlnProArgProGlyThrThr-----SerGlnAlaSer	731
OY	129	TGTTGTATGATGAGCCACAGACGGGTATGCTTCTCTGGG	93

```

Db          |||      ::|||:::   |||||
731 rGlyProclYasnSerSerThrsrThrlyGProGly 743

RESULT 9
PCT-US95-04611A-19
; Sequence 19, Application PC/TUS9504611A
; GENERAL INFORMATION:
APPLICANT: Spaete, Richard and Jackman, Wintthrop, T
TITLE OF INVENTION: Non Splicing Variants of gp350/220
NUMBER OF SEQUENCES: 19
CORRESPONDENCE ADDRESS:
ADDRESSEE: Cooley Godward Castro Huddleson & Tatum
STREET: 5 Palo Alto Square
CITY: Palo Alto
STATE: California
COUNTRY: USA
ZIP: 94306
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC Compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US95/04611A
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/229,291
FILING DATE: April 18, 1994
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Luann Cseri
REGISTRATION NUMBER: 31,822
REFERENCE/DOCKET NUMBER: AVIR-003/00US
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415-843-5163
TELEX: 415-857-0663
FAX: 380816 CoolEYPa
INFORMATION FOR SEQ ID NO: 19:
SEQUENCE CHARACTERISTICS:
LENGTH: 907 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
PCT-US95-04611A-19

Alignment Scores:
Pred. No.: 0.00378 Length: 907
Score: 115.00 Matches: 90
Percent Similarity: 32.77% Conservative: 46
Best Local Similarity: 21.69% Mismatches: 117
Query Match: 4.49% Indels: 162
DB: 5 Gaps: 18

US-09-989-919-15 (1-1397) x PCT-US95-04611A-19 (1-907)

QY 1182 CCCACCCCGGCCAGGT---CTGGCTCAAACCAACC----- 1147
||| ||| ||| |||
Db 442 ProsnhnrhrrhnglyleuProserSerthrhlstvalProthraenleuthralaPro 461
|||||
QY 1146 TGCTTCCTGACCCCAAGTCTTGCTCATGACTGAGGTGCACCTTGCCCTGACCCCTGG 1087
||| ||| ||| |||
Db 462 AlaserThrglyPromhValserThrlalaspyal-----ThsrPro----- 476
|||||
QY 1086 GCTGGCCTGGAGGTGCTCCCTGGTGGGAGGTCAGAGATAGCTCCCAAGTACGAATCA 1027
|||||
Db 477 -----ThPralaGlyThrThrsrGlYAlasrProvalThrProser 491
|||||
QY 1026 CCCACATCTGAGACATCCCGGCCAAGTCTCTGTCACAGACTTAAGCTGAGAGAACA 967
|||||
Db 492 ProserProTrlpapasmnglyThnGluseryslalProbapmetThrsrThsr 511
|||||
QY 966 CCACAGTAGATTGTTCCCAAGCACGACCCCAAGCCGCTGCTGTGGCATGTGACGACAGCA 907

```

Db 512 Pro-----valthrthrProthrProasnaIahrsrProthrProalaVal 527
 QY 906 GCGGGGGCTTACTTACTTCTACACTATTGAACAGATGGGGCTCCACGGAAC 847
 Db 528 Thr----- 528
 QY 846 CGTCCAGAGGAGGGGACAGCCCCAGTGGGAGACTAGCCTGAGCTTGCTTACAGCCCA 787
 Db 529 -----ThrPro 530
 QY 786 CAGGGTTGAGAGGACGATACCCCATGATT----- 754
 Db 531 ThrProasnaIahrsrProthrProalaValthrThrProthrProasnaIahrsr 550
 QY 753 CCGGACTAGAGGACGACGATCCCATAGCTAGAGCTGTAACTACATTTTACATTGG 654
 Db 551 ProthrleuGIythrThrSerProthrSerAlaValthrThrProthrProasn----- 568
 QY 693 CATTATTGACGTTTGCCTGCTGATCTGCTGAGATGTTAGCTTTGCTGACG 634
 Db 569 -----AlathrSerProthrleuGIy-LysThr-----SerPr 579
 QY 633 GGGTCCAGAGTACATGACATGCTGCTGACGATGCAAGTACTCCCTCCACAGGAG----- 578
 Db 579 othrSerAlaValthrThrProthrProasnaIahrsrProthrleuGIyLysThr 599
 QY 577 -----CCAGCTGCTAGAGTACTGCTGCTGACG 547
 Db 599 erProthrSerAlaValthrThrProthrProasnaIahrsrProthrValGIyLthr 619
 QY 546 CGAGCCCTGAG----- 536
 Db 619 hrSerProGlnAlaAsnaIahrsrHsrThrleuGIyLthrSerProthrProValy 639
 QY 535 --ATAGGTCCAGGAGGCTTCTGATGCTGTTGGATGAGGACCATTTGGC 478
 Db 639 althrSerGlnProlyAsnaIahrsrAlaVal-----ThrThrGIy 654
 QY 477 TCATCCAGCTTGAATAAAC-----AGCTCCGACCTGGGGGAGAGGT 430
 Db 654 InhrAsnaIahrsrSerSerThrSerSerMetSerLeuArgProSer----- 670
 QY 429 GAGAGGTAACAGGAGGCTGATGTTCTTCTGAGGACAGACCATCTCCCTAATTGA 370
 Db 671 -----SerasnProGlnThrLeuSerProSerThr 681
 QY 369 CTGCTTGAAGACGATACCTCTCTCATCACTCAGACCCCTCTGACTTCGAT 310
 Db 681 erAspanserThrSer-HisMetProleuThrSerAlahsrPro-----Thr 697
 QY 309 GGGGATCACTATCTATCAACCTCCAGAGGCTCCCATCTCCCTTGGAGAGGCTTCT 250
 Db 698 GylGIyGIy-----AsnIle-ThrGlnValthrProIahsrIle----- 710
 QY 249 CCACATGGCTCACCTGTCAGAGGTGTGTGGTGTACCAAAAGCCGACATGGGCA 190
 Db 711 -----SerThrHisI 714
 QY 189 TGGCTCTACAGACATCCACAGCTCAGCAGGTTGAACATGAAGAGAGAGCTCCG 130
 Db 714 sValserThrSerSerProGlnProArgProGIyLthrThr-----SerGlnIahsr 731
 QY 129 TGGTGTGAGATGGCCAGCAGCGGTAGCTTCTCTGGG 93
 Db 731 rGIyProGIyAsnSerSerThrSerThrLysProGIy 743

RESULT 10
 US-08-127-499A-8
 Sequence 8, Application US/08127499A
 Patent No. 5510264
 GENERAL INFORMATION:
 APPLICANT: VAN ALSTYNE, Diane

APPLICANT: SHARMA, Lawrence Rajendra
 TITLE OF INVENTION: ANTIBODIES WHICH BIND MENINGITIS RELATED
 TITLE OF INVENTION: HOMOLOGOUS ANTIGENIC SEQUENCES
 NUMBER OF SEQUENCES: 40
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Foley & Lardner
 STREET: 3000 K Street, N.W., Suite 500
 CITY: Washington
 STATE: D.C.
 COUNTRY: USA
 ZIP: 20007-5109
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: Patent Release #1.0, Version #1.30
 CURRENT APPLICATION DATA:
 FILING DATE: 28-SEP-1993
 APPLICATION NUMBER: US/08/127,499A
 ATTORNEY/AGENT INFORMATION:
 NAME: BENT, Stephen A.
 REGISTRATION NUMBER: 29,768
 REFERENCE/DOCKET NUMBER: 51916/102/INBI
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (202) 672-5300
 TELEFAX: (202) 672-5399
 TELEX: 904136
 INFORMATION FOR SEQ ID NO: 8:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 1063 amino acids
 TYPE: amino acid
 STRANDEDNESS:
 TOPOLOGY: unknown
 US-08-127-499A-8
 Alignment Scores:
 Pred. No.: 0.00406 Length: 1063
 Score: 115.00 Matches: 95
 Percent Similarity: 31.65% Conservative: 37
 Best Local Similarity: 22.78% Mismatches: 116
 Query Match: 4.49% Indels: 149
 Gaps: 20
 US-09-989-919-15 (1-1397) x US-08-127-499A-8 (1-1063)
 QY 1184 GCGCCACCCCGCCGAGTCTGCTCAACC-----ACACTGCTCCCTGACCC 1134
 Db 77 AlaProProProGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 96
 QY 1133 CAGTCTGCTCATGTGTGAGTGTGACCTGACCTGACCCCTGAGCGCTGGCTGGAG 1074
 Db 97 SerArg-----AlaPro-----ProGlnGlnProGln----- 105
 QY 1073 TGTCCCTGTCGGAGGTGAGATAGCTCCCGACAGTACAG-----AAT 1029
 Db 106 -----ProArgMetGlnThrGlyArgGlyGlySer 116
 QY 1028 CACCCACATCTGAGACATCCCGCCCAAGTCTCTGCTCAGACCTTAAGTGAGAA 972
 Db 117 AlaProArgProGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 136
 QY 971 -----ACACTCCAGATGATGTGCCACGACGAGCCCGCCAGCGCTGCTG 926
 Db 137 LeuArgProProleuHisAspPro-----AspThrGln-AlaProThrGlnAlaCysVal 155
 QY 925 -----GGCATGTCAGACGACGACGAGGAGGCTTACTTCACTTCTCAGACTTAAGAAC 870
 Db 155 rSerThrProleuTrpSerGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 175
 QY 869 AGTATGGGGCTGCGCCACGGAACCTGCCAGGAA----- 836
 Db 175 rAenleuGIyThrProProleuAspGlnAspGlyArgTrpAspProAlaLeuMetLys 195

QY 835 -----GGGGCAGAGCCCGAGTGGGAGACTAGACCTTGCTTACAGCCACAGG 783
 Db 195 nProCysGlyProGlnProPro---AlaHisValValArgAlaIstyrAsnGlnProIleagl 214
 QY 782 GTTGGAGAGGACGATCATAACCCCATGATTCCTGGACATAGAGCAGAGATCCCATAGCTA 723
 Db 214 yAspValArgGlyValIstyrGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGly 234
 QY 722 GGCTGGTAACATCATTTTAACTTTGACATTTGACATTTGACATTTGACATTTGACATTTG 663
 Db 234 IGlyGly-----ThiArgTrpHisArgLeuLeu-----ArgMetPro 246
 QY 662 TCTGTCTGGAGATGTGATCTTGTCTGAGAGGAGTTCAGATGACATGACATGACATGACAT 603
 Db 246 ovalArgGlyLeuAspIstyrAspSerAlaProLeuProPro---HisThrThrGlyIst 265
 QY 602 GCAAGTCATCTCCCTCCACAGCAGCAGCCAGCTGTGAGATGATGACATGACATGACATG 545
 Db 265 eGluThrArg-----SerIleArgHisProTr 274
 QY 544 -AGCCCTCAGATAGTGCAGCAGAGGCTTCACATATGTGCTGTGTGGATCAGAGGAC 486
 Db 274 pArgIleArgPheGlyAlaProGlnAlaPheLeuAlaGlyLeuLeuLeuAlaThrValAl 294
 QY 485 CACTGTGCTCCATCCAGCTCTTGAAGAGAGCAGCTGCCAGCTGGGGGAGCAGAGTGAGA 426
 Db 294 a-ValGlyThrAlaArgAlaGlyLeuGlnProAlaArgAlaAspMet----- 308
 QY 425 GGTACAGAGGCTGATTTCTCTTTCTGGAGGAGACACATTCCTCCATATTGACTGC 366
 Db 308 ----- 308
 QY 365 CTGAGAGACAGATATACCTCTCTCTATCACTCAGCAGCCCTTGACTTCGAGTGGG 306
 Db 309 -----AlaAlaProProIleProLeuProGlnPro 317
 QY 305 GATCACTATCTATCAACTCCAGAGAGCTCCCATCTCCCTTGAGAGGAGCTTCTCCCA 246
 Db 318 -----Pro 318
 QY 245 CTGGCTCAGCTGTCAGAGTGTCTGTGTGTGACACAAAGCCCGACACTGGGGCATGGC 186
 Db 319 CysAlaHisIstyrGlnIstyrGlyHisHisHisHisIstyrProPheLeuGlyHisAsp 338
 QY 185 TCTCAGACATCCAGAGCTCAGAGGAGTGAACAATGAAGAGGAGCAGCTCCCGGT 126
 Db 339 GlnHisHis-----GlyGlyThrLeuArgVal 347
 QY 125 GGTAGATGGCCAGAGCGGTAGT-----CTTCTGGGGGATGG 87
 Db 348 GlnGlnHisIstyrArgAsnAlaSerValLeuProGlyHisIstyrLeuGlnGlyGlyTrp 367
 QY 86 TGCTGTCTGGGA-----TACACTGTGACTCGTACTGC 54
 Db 368 GlnCysIstyrAsnLeuSerAspTrpHisGlnGlyThrHisValCys 382
 RESULT 11
 US-08-482-847-8
 Sequence 8, Application US/08482847
 Patent No. 5556757
 GENERAL INFORMATION:
 APPLICANT: VAN ALSTYNE, Diane
 APPLICANT: SHARMA, Lawrence Rajendra
 TITLE OF INVENTION: PEPTIDES REPRESENTING EPITOPIC SITES FOR
 TITLE OF INVENTION: BACTERIAL AND VIRAL MENINGITIS CAUSING AGENTS AND THEIR
 TITLE OF INVENTION: CNS CARRIER, ANTIBODIES THEREOF, AND USES THEREOF
 NUMBER OF SEQUENCES: 40
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Foley & Lardner
 STREET: 3000 K Street, N.W., Suite 500
 CITY: Washington
 STATE: D.C.
 COUNTRY: USA

ZIP: 20007-5109
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 OPERATING SYSTEM: IBM PC compatible
 SOFTWARE: Patent in Release #1.0, Version #1.30
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/482,847
 FILING DATE: 07-JUN-1995
 CLASSIFICATION: 514
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 08/127,499
 FILING DATE: 28-SEP-1993
 ATTORNEY/AGENT INFORMATION:
 NAME: BENT, Stephen A.
 REGISTRATION NUMBER: 29,768
 REFERENCE/DOCKET NUMBER: 51916/104/INBI
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (202) 672-5300
 TELEFAX: (202) 672-5399
 TELEX: 904136
 INFORMATION FOR SEQ. ID NO.: 8:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 1063 amino acids
 TYPE: amino acid
 STRANDEDNESS:
 TOPOLOGY: unknown
 US-08-482-847-8
 Alignment Scores:
 Pred. No.: 0.00406 Length: 1063
 Score: 115.00 Matches: 95
 Percent Similarity: 31.65% Conservative: 37
 Best Local Similarity: 22.78% Mismatches: 136
 Query Match: 4.49% Indels: 149
 Gaps: 20
 US-09-989-919-15 (1-1397) x US-08-482-847-8 (1-1063)
 QY 1184 GCGCCAGCCCGCCGAGGCTCCGCTCAAC-----ACACTGTCTCCGACCC 1134
 Db 77 AlaProProProIleGlnArgGlnGlnSerArgSerGlnThrProAlaProLysPro 96
 QY 1133 CAGCTTGCTCATGCTGAGTGTGACCTTCGCTGACCTGACCTGAGGCTGGCTGGAG 1074
 Db 97 SerArg-----AlaPro-----ProGlnGlnProGln----- 105
 QY 1073 TGTCCCTGTGCGGAGGTGAGAGATAGCTCCCGACGATAG-----AAT 1029
 Db 106 -----ProProArgMetGlnThrGlyArgGlySer 116
 QY 1028 CACCCACATCTCGAGACATCCCGCCAAAGTCTCTGTCCAGACCTTAAGCTGAGAG 972
 Db 117 AlaProArgProGlnLeuGlyProProThrAsnProPheGlnAlaAlaValAlaArgGly 136
 QY 971 -----ACACTCAGATGATGTGCTCCCAAGCAGCCCGACGCTGCTG----- 926
 Db 137 LeuArgProProLeuHisAspPro-----AspThrGln-AlaProThrGlnAlaCysValIle 155
 QY 925 -----GGCATGTGACAGCAGCAGAGGGGCTTTAGCTATCTACACTATTGAACAC 870
 Db 155 rSerTrpLeuTrpSerGlnGlyGlnGlyAlaValAlaPheTrpArgValAspLeuHisPhe 175
 QY 869 AGTATGAGGCTGCGCCAGAGGACCGCTCCAGGAG----- 836
 Db 175 rAsnLeuGlyIleProProLeuAspGlnAspGlyArgTrpAspProAlaLeuMetTrpAs 195
 QY 835 -----GGGGCAGAGCCCGAGTGGGAGAGCTGAGCTGTACAGCCACAGG 783
 Db 195 nProCysGlyProGlnProPro---AlaHisValValArgAlaIstyrAsnGlnProIleagl 214
 QY 782 GTTGGAGAGGACGATCATAACCCCATGATTCCTGGACATAGAGCAGAGATCCCATAGCTA 723

```

Db 214 yAspValArgGlyValThrGlyLysGlyGluArgThrTyAlaGluGlnAspPheArgVa 234
QY 722 GGCTGGAAATCACTATTAACTTTGGCATTTCACAGCTTTGCTGCTGACATCTG 663
Db 234 IglyGly-----ThrArgTrpHisArgLeuLeu-----ArgMetPr 246
QY 662 TCGTCTGGAGATAGCTTGTCTGTGACAGGGGTCCAGAGTACATGACATGCGCCAGT 603
Db 246 oValArgGlyLeuAspGlyAspSerAlaProLeuProPro---HisThrThrGluArgI 265
QY 602 GCAAGTACTCCCTCCACAGGCAAGCCAGCTGCTGAGATAGTACAGTGGTCCAGCCG-- 545
Db 265 egluThrArg-----SerAlaArgHisProTyr 274
QY 544 -AGCCCTCAGATAGTACAGCCAGAGGCTTCACATATGTGTCTTTGGATGACAGGAC 486
Db 274 pArgileArgPheGlyAlaProGlnAlaPheLeuAlaGlyLeuLeuLeuAlaThrValAl 294
QY 485 CACTGTGCTCCATCAGCTTTGAAAGAACAGCAGCTGCGGAGCTGGGGGAGAGGTGAGA 426
Db 294 a-ValGlyThrAlaArgAlaGlyLeuGlnProAlaAlaAspMet----- 308
QY 425 GGTAAAGGCTGATGTTCTCTTTCTGAGGCAAGCATTCTCCCTAATTGACTGC 366
Db 308 ----- 368
QY 365 CTGAAGACAGATACCTCTCTCTCATCAGCTCAGACCCCTCTGATTCGGATGGG 306
Db 309 -----AlaAlaProProThr-LeuProGlnPro 317
QY 305 GATCATATCTATCAACCTCCAGAGAGTCTCGCATCTCCCTGAAAGGCTTCTCCCA 246
Db 318 -----Pro 318
QY 245 CTGAGTCACTGTCCAGAGTGTCTGTGTGTGACCAAGGCCGCACTGGGATGGC 186
Db 319 CysAlaHisIstYrGlnHisIstYrGlyHisIstHisIstLeuProPheLeuGlyHisAsp 338
QY 185 TCTCCACAGATCCACAGCTCCAGGCTGTTGAACACTGAAGAGGACACTCCCGTGT 126
Db 339 GlyHisHisIst-----GlyGlyThrLeuArgVal 347
QY 125 GGTAGATGGCCAGCAGCGGTAGT-----CTTCCGGGGAGTGG 87
Db 348 GlyGlnHisIstYrArgAlaAspValLeuProGlyHisIstYrLeuGlnGlyGlyTyr 367
QY 86 TGCTGTCTGGA-----TACACTGTGACTGGTACTGC 54
Db 368 GlyCysIstYrAsnLeuSerAspTrpHisGlnGlyThrHisValCys 382

```

```

RESULT 12
US-08-874-569B-20
; Sequence 20, Application US/08874569B
; Patent No. 6306650
; GENERAL INFORMATION:
; APPLICANT: Townes, Tim M.
; APPLICANT: Donze, David
; TITLE OF INVENTION: DELTA-BRYTHROID KRUPPEL-LIKE FACTORS AND
; TITLE OF INVENTION: METHODS OF USE
; FILE REFERENCE: 05118.000802
; CURRENT APPLICATION NUMBER: US/08/874,569B
; CURRENT FILING DATE: 1997-06-13
; PRIOR APPLICATION NUMBER: 60/019,769
; PRIOR FILING DATE: 1996-06-14
; NUMBER OF SEQ ID NOS: 21
; SOFTWARE: FastSeq for Windows Version 4.0.
; SEQ ID NO 20
; LENGTH: 376
; TYPE: PRT
; ORGANISM: Murine Beta-EKLF
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence:\ No. 6306650e =
; OTHER INFORMATION: synthetic construct

```

```

US-08-874-569B-20
Alignment Scores:
Pred. No.: 0.00464 Length: 376
Score: 112.50 Matches: 80
Percent Similarity: 32.23% Conservative: 27
Best Local Similarity: 24.10% Mismatches: 97
Query Match: 4.39% Indels: 128
DB: 4 Gaps: 18

US-09-989-919-15 (1-1397) x US-08-874-569B-20 (1-376)
QY 1179 ACCCCGCCCCAGGT-----CCTGGCTAAACACACCTGTCCCTGACCCAGTCC 1129
Db 98 ThrAspPheProGlySerLeuSerProGlyThrSerArgThrCysAlaLeuAlaProSer 117
QY 1128 TTGGCTCATGTCTGAGGTGTGACCTTGTCCCTGACCCCTGGGCTGGCTGGAGTGTCC 1069
Db 118 ValGlyPro-----ValAlaGlnPheGlnProProGlnSerLeuGlyAlaTyr 133
QY 1068 CCTGTGGAGGTGACAGATAGCTCCCA---GGTACAGATCACCACATCTGTGAGC 1012
Db 134 AlaGlyGlyProGlyLeuValThrGlyProLeuGlySerGlnGlyHisIstHisSerTrpAla 153
QY 1011 ATCCCGGCGAAGTCTCTGTCCAGACTAAGCTGAGAGAACACTCCACAGATGATGTC 952
Db 154 HisPro-----ThrProArg----- 158
QY 951 CCCAGACAGCCCCACCGCTGTGTGAGATGTGACAGCAGCAGCAGGAGGCTTAACT 892
Db 159 -----ProProAlaProGlnProPheValAlaProAlaLeuAlaProGlyLeuAla 175
QY 891 ACATTCTACACTATTGAACACAGATAGGAGCTGCCACGAGACCGTCCAG----- 838
Db 176 -----ProLysAlaGlnProSerTyrSer 183
QY 837 -----GAGGGACAGCCCCAGTGGGA----- 814
Db 184 AspSerArgAlaGlySerValGlyGlyPhePheProArgAlaGlyLeuAlaValProAla 203
QY 813 -----GACTAGCTAGAGTGTCTTACAGCCACAG 784
Db 204 AlaProGlyAlaProTyrGlyLeuLeuSerGlyTyrProAlaLeuTyrProAlaProGln 223
QY 783 GGTGAGAGGACAGATATACCCCATGATCTCTGACTGTGAGCAGCAGCAGTCCATAGCT 724
Db 224 ---TyrGlnGly---HisPheGlnLeuPheArgGlyLeuAlaAlaProSerAlaGlyGly 241
QY 723 AGGCTGTAACTCAGATTTTAACTATTGACATTATGACAGTTTCTGTCGACATCT 664
Db 242 ThrAlaProProSerPheLeuAsn----- 249
QY 663 GTCTGTCTGGATGTAGTTGTTGTGACAGGGTTCCAGGTGACATGACAGTGTCCAG 604
Db 250 ---CysLeuGly-----ProGly-ThrValAlaThrGluLe 260
QY 603 TGCAGTACTCCCTCCACAGCAGCAGCCAGCTGTGAGAGATAGTACGTGTGACCCGA 544
Db 260 uGlyAlaThrAlaIleAlaGlyAspAlaGlyLeu-----SerProGlyThrAl 276
QY 543 GCCCTAGATAGTACAGCAGAGGCTT-----CACATATGTG 505
Db 276 aProProLysArgSerArgArgThrLeuAlaProLysArgGlnAlaHisIstHisSerG 296
QY 504 TCTTGTGGG-----ATCAGGAGCACTTGCTCC 475
Db 296 yHisGlnGlyCysGlyLysSerTyrSerLysSerHisLeuLysAlaHisIstLeuArgTh 316
QY 474 ATCCAGTCTTGAAGAACAGCAGCTGCG---GGGACAGAGTGA--- 426
Db 316 rHisThrGlyGlyLysProTyrAlaCysSerTrpAspGlyCysAspTrpArgPheAlaAr 336
QY 425 -----GTTAACAGGCTGAGATGTTCT-- 405

```

Db 336 gSerAspGluLeuThrArgHisTyrArgLysHisThrGlyHisArgProHeCysCysGI 356
 QY 404 -----TCTTTCGAGAGCAAGACCAT 384
 Db 356 yLeuCySProArgAlaPheSerHisgSerAspHis 367

RESULT 13

US-08-127-499A-1

Sequence 1, Application US/08127499A

Patent No. 5510264

GENERAL INFORMATION:

APPLICANT: VAN ALSTYNE, Diane

APPLICANT: SHAMA, Lawrence Rajendra

TITLE OF INVENTION: ANTIBODIES WHICH BIND MENINGITIS RELATED

TITLE OF INVENTION: HOMOLOGOUS ANTIGENIC SEQUENCES

NUMBER OF SEQUENCES: 40

CORRESPONDENCE ADDRESS:

ADDRESSEE: Foley & Lardner

STREET: 3000 K Street, N.W., Suite 500

CITY: Washington

STATE: D.C.

COUNTRY: USA

ZIP: 20007-5109

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patent in Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/127,499A

FILING DATE: 28-SEP-1993

ATTORNEY/AGENT INFORMATION:

NAME: BENT, Stephen A.

REGISTRATION NUMBER: 29,768

REFERENCE/DOCKET NUMBER: 51916/102/INBI

TELEPHONE: (202)672-5300

TELEFAX: (202)672-5399

TELEX: 904136

INFORMATION FOR SRO ID NO: 1:

SEQUENCE CHARACTERISTICS:

LENGTH: 992 amino acids

TYPE: amino acid

STRANDEDNESS:

TOPOLOGY: unknown

US-08-127-499A-1

Alignment Scores:

Pred. No.: 0.008 Length: 992
 Score: 112.00 Matches: 97
 Percent Similarity: 31.9% Conservative: 38
 Best Local Similarity: 22.9% Mismatches: 129
 Query Match: 4.38% Indels: 159
 DB: 1 Gaps: 22

US-09-989-919-15 (1-1397) x US-08-127-499A-1 (1-992)

QY 1184 GCGCCACCCCGCCGCTGCTGCTCAACC-----ACACTGCTCCCTGACCC 1134
 Db 76 AlaProProProProGluArgGlnGlnSerArgSerGlnThrProAlaProLysPro 95
 QY 1133 CAGCTTGCTCATGCTGAGAGTGTGACCTGTGCGCCCTGACCCCTGGCGCTGGAG 1074
 Db 96 SerArg-----AlaPro-----ProGlnGlnProGln----- 104
 QY 1073 TGTCCCTGTGGAGGTGAGATAGCTCCCGAGGTACAG-----AAT 1029
 Db 105 -----ProProArgGlnGlnThrArgLysGlySer 115
 QY 1028 CACCCACATCTGTGAGATCCCGCCCAAGTCTCTTCACAGACTTAAGTGAAGA--- 972
 Db 116 AlaProArgProGluLeuGlyProProThrAsnProHeGlnAlaAlaValAlaArgGly 135

QY 971 -----ACATCCAGATGATGTCGCCAGCGAGCCCGCCGCTGTCT----- 926
 Db 136 LeuArgProProLeuHisAspPro-----AspThrGlu-AlaProThrGluAlaCysValTh 154
 QY 925 ---GGCATGTGACGACGACGACGAGGGGCTTTAGCTACATTTCACTATTGAAC 870
 Db 154 rSerThrLeuTrpSerGlnGlyGluGlyAlaValPheTyrArgValAspLeuHisPhe 174
 QY 869 AGTGAATGGGGCTGCGCCAGGAAACCGCCAGGGA----- 836
 Db 174 eAsnLeuGlyThrProProLeuAspGluAspGlyArgTyrPheProAlaLeuMetTyrAs 194
 QY 835 -----GGGACACAGCCCGCCAGTGGGAGAGACTAGCTGCTGCTTACAGCCAG 783
 Db 194 nProCysGlyProGluProPro-----AlaHisValAlaValAlaTyrAsnGlnProAlaG 213
 QY 782 GTTGAGAGGACAGTCAATACCCCATGATTCCTGAGCTAGAGCCAGCTCCCATAGCTA 723
 Db 213 yAspValArg-----GlyValTrpGlyGlyGluArgThrTyr-Ala- 227
 QY 722 GCGGTAAACTCACTTTTAACAT-----TGCAATTATGCACTTTGTCC 675
 Db 228 -----GluGlnAspPheArgValGlyGlyThrArgThrPheArgLeuLeuArgMetP 245
 QY 674 TGCTCACATCTGTCTGTGGAGTGTAGCTTTGTCTGAGAGGGCTTCCAGGTACATGC 615
 Db 245 roVal-----ArgGlyLeuAspGlyAsp----- 252
 QY 614 AGTGCTGCAGTGCAGTCACTCTCTCCACAGGCAAGCCAGCTGTGAGATAG-TCA 556
 Db 253 -----ThrAlaProLeuProProHisThrThrGluArgGlnThrArgSerA 269
 QY 555 CTGCTCAGCG---ACCCTCAGATAGTGCAGCAGGAGCTTCAATATGTGCTGTG 499
 Db 269 laArgHisProTrpArgGlnLeuArgPheGlyAlaProGlnAlaPheLeuAlaGlyLeuLeu 289
 QY 498 TGGGATCAGGAGCACTTGTGCTCCATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 439
 Db 289 eulaAlaValAla-ValGlyThrAlaArgAlaGlyLeuGlnProArgAlaAspMet-- 307
 QY 438 GCGAGAGTGAAGAGTAACAGGCTGGATGTTCTCTTCTGAGGCAAGACCATTTCTC 379
 Db 307 ----- 307
 QY 378 CTAAATTGACTGCTTGAAGACAGCTATACCTCTCTCATACCTCAGACCCCTCTG 319
 Db 308 -----AlaAlaProPheMet 312
 QY 313 -----Pro-ProGlnPro----- 316
 QY 258 AGGGCTTTCCTCCACTGCTCAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 199
 Db 317 -----ProArgAlaHisGlyGlnHisTyrGlnHisHisGlnLeuProH 333
 QY 198 CACTGGCATGGCTCTCAGACATCAGACCTCAGACGCTGTAACACTGAAGAGG 139
 Db 333 eLeuGlyHisAspGlyHisHis-----GlyG 342
 QY 138 CAGCTCCGCTGGGTGAGAGTGGCCAGAGCGGTACT-----CT 100
 Db 342 yThrLeuArgValGlyGlnHisHisArgAsnAlaSerAspValLeuProGlyHisTyrPhe 352
 QY 99 TCCTGGGAGATGCTGTCTGAGGA-----TACACTGTACTCTGACTGTC 54
 Db 362 uGlnGlyGlyTrpGlyCysTyrAsnLeuSerAspTrpHisGlnGlyThrHisValCys 381

RESULT 14

US-08-482-847-1
 Sequence 1, Application US/08482847
 Patent No. 5556757

GENERAL INFORMATION:

APPLICANT: VAN ALSTINE, Diane
 APPLICANT: SHARMA, Lawrence Rajendra
 TITLE OF INVENTION: PEPTIDES REPRESENTING EPITOPIC SITES FOR
 TITLE OF INVENTION: BACTERIAL AND VIRAL MENINGITIS CAUSING AGENTS AND THEIR
 TITLE OF INVENTION: CNS CARRIER, ANTIBODIES THERETO, AND USES THEREOF
 NUMBER OF SEQUENCES: 40
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Foley & Lardner
 STREET: 3000 K Street, N.W., Suite 500
 CITY: Washington
 STATE: D.C.
 COUNTRY: USA
 ZIP: 20007-5109
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: Patent in Release #1.0, Version #1.30
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/482,847
 FILING DATE: 07-JUN-1995
 CLASSIFICATION: 514
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 08/127,499
 FILING DATE: 28-SEP-1993
 ATTORNEY/AGENT INFORMATION:
 NAME: BENT, Stephen A.
 REGISTRATION NUMBER: 29,768
 REFERENCE/DOCKET NUMBER: 51916/104/INBI
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (202)672-5300
 TELEFAX: (202)672-5399
 TELEX: 904136
 INFORMATION FOR SEQ ID NO: 1:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 992 amino acids
 TYPE: amino acid
 STRANDEDNESS:
 TOPOLOGY: unknown
 US-08-482-847-1

Alignment Scores:

Pred. No.: 0.008 Length: 992
 Score: 112.00 Matches: 97
 Percent Similarity: 31.99% Conservative: 38
 Best Local Similarity: 22.99% Mismatches: 129
 Query Match: 4.38% Indels: 159
 Gaps: 22

US-09-989-919-15 (1-1397) X US-08-482-847-1 (1-992)

QY 1184 GCGCCACCCCGCCGAGTCCGCTGCAAC-----ACACCTGCTCCCTGACCC 1134
 DB 76 AlaProProProProGluGluArgGlnGlnSerArgSerGlnThrProAlaProLysPro 95
 QY 1133 CAGCTTGCGCTCAIGCTAGGTGACAGCTGCTGCGCCCGACCCCTGGGCTGCGGGAG 1074
 DB 96 SerArg-----AlaPro--ProGlnGlnProGln----- 104
 QY 1073 TGTCCCTGTGGGAGGTCAAGATAGCTCCCGCAGGTACG-----AAT 1029
 DB 105 -----ProProArgMetGlnThrGlnArgGlyGlySer 115
 QY 1028 CACCCACATCTTGAGCATCCCGCAGGTCTCTGTCAGACCTAAGTGAAGA-- 972
 DB 116 AlaProArgProGluLeuGlyProProThrAsnProPheGlnAlaAlaValAlaArgGly 135
 QY 971 -----ACACTCCAGATGATGTTGCCACCGCAGCCCGCGCTGCTGT----- 926
 DB 136 LeuArgProProLeuHisAspPro--AspThrGln-AlaProThrGlnAlaCysValTh 154
 QY 925 ---GGCATGTGCACAGCAGACAGAGGGGCTTAGTACATTCACACTATTGAACAC 870

DB 154 rSerTrpLeuTrpSerGlnGlyGlnAlaValPheThrArgValAspLeuHisPhe1 174
 QY 869 AGTATGGGGCTCCCGCAGGAAACCGTCCAGGA----- 836
 DB 174 eAsnLeuGlyThrProProLeuAspGluAspGlyArgTrpAspProAlaLeuMetTyrAs 194
 QY 835 -----GGGGACAGCCCGCCAGTGGGAGACTGAGCCCTGCTTACAGCCACAGG 783
 DB 194 nProCysGlyProGluProPro-----AlaHisValValArgAlaThrAsnGlnProAlaG1 213
 QY 782 GTTGAAGAGCAGTCATACCCCATGATTCCTGCACTAGAGCCAGCAGCTCCATAGCTA 723
 DB 213 yAspValArg-----GlyValThrPheGlyGlnArgGlnThrTyr-Ala- 227
 QY 722 GCGCTGTAACATCATTTTAACAT-----TGGCATTTATGACAGTTTGGC 675
 DB 228 -----GlnGlnAspPheArgValGlyGlyThrArgTrpHisArgLeuLeuArgMetP 245
 QY 674 TGTCACATCTGTCTGTGGAGTGTAGCTTTGCTGCAGGGGTTCCAGGTGACATGC 615
 DB 245 roval-----ArgGlyLeuAspGlyAsp----- 252
 QY 614 AGTGTGCCAGTGAAGTCACTCCCTCCACAGGAAAGCCAGCTGTGAGATAG-TAG 556
 DB 253 -----ThrAlaProLeuProProHisThrThrGlnArgGlnThrArgSerA 269
 QY 555 CTGTGACAGCG--AGCCCTGAGATAGGTGACAGCCAGGCTTCATATGAGTGTGT 499
 DB 269 lArgHisProTrpArgIleArgPheGlyAlaProGlnAlaPheLeuAlaGlyLeuLeu 289
 QY 498 TGGGATCAGGAGCAGCTGCTCCATCCAGTCTTGAAGAAAGCAGCTGCCAGCTGGGG 439
 DB 289 eAlaAlaValAla-ValGlyThrAlaArgAlaGlyLeuGlnProArgAlaAspMet--- 307
 QY 438 GCGAGAGTGAAGGTACAGGCGTGGATGTTCTTTCTTTCGAGGCAAGACCATTTCC 379
 DB 307 ----- 307
 QY 378 CTAAATTGACCTGCTTGAAGACAGGTATACCTCTCTCATCAGCTCAGACCCCTGTG 319
 DB 308 -----AlaAlaProPromet 312
 QY 318 ACTTCGATGGGGATCACTATCTATCAACTCCAGAGAGTCTGGCATCTCCCTTGA 259
 DB 313 -----Pro-ProGlnPro----- 316
 QY 258 AGGGCTTTCACATGCTGCTCAGCTGTCAGAGTGTGTGTGTGACCCAGAGCCGGA 199
 DB 317 -----ProArgAlaHisGlyGlnHisTyrGlnHisHisHisGlnLeuProPh 333
 QY 198 CACTGGGATGCTGTCCACAGACATCCACAGCCTCAGCAGGTGAAACCTGAAAGAGG 139
 DB 333 eLeuGlnHisAspGlyHisHis-----GlyG1 342
 QY 138 CAGCTCCCGTGTGTAGGATGGCCAGCAGCGTAGT-----CT 100
 DB 342 yThrLeuArgValGlyGlnHisHisArgHisAspValLeuProGlyHisTyrPhe 362
 QY 99 TCTGAGGAGATGCTGTCTGGA-----TACACTGTGACTCGGACTGC 54
 DB 362 uGlnGlyGlyTrpGlyCysTyrAsnLeuSerAspTrpHisGlnGlyThrHisValCys 381

RESULT 15

US-08-690-473-2
 Sequence 2, Application US/08690473

Patent No. 5876923

GENERAL INFORMATION:

APPLICANT: Leopold, Rosario
 APPLICANT: Roizman, Bernard
 TITLE OF INVENTION: HERPES SIMPLEX VIRUS ICp4 AS AN
 TITLE OF INVENTION: INHIBITOR OF APOPTOSIS
 NUMBER OF SEQUENCES: 2


```

CORRESPONDENCE ADDRESS:
ADDRESSEE: Arnold, White & Durkee
STREET: P.O. Box 4433
CITY: Houston
STATE: Texas
COUNTRY: USA
ZIP: 77210

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/690,473
FILING DATE: 26-JUL-1996
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Highlander, Steven L.
REGISTRATION NUMBER: 37,642
REFERENCE/DOCKET NUMBER: ARCD:239
TELECOMMUNICATION INFORMATION:
TELEPHONE: 512/418-3000
TELEFAX: 512/474-7577
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 1298 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: linear
US-08-690-473-2

Alignment Scores:
Pred. No.: 0.0162 Length: 1298
Score: 109.50 Matches: 102
Percent Similarity: 28.81% Conservative: 19
Best Local Similarity: 24.29% Mismatches: 131
Query Match: 4.28% Indels: 168
DB: Gaps: 22

US-09-989-919-15 (1-1397) x US-08-690-473-2 (1-1298)
QY 1143 TCCTGACCCGAGTCTTGCTGCTAGTGAAGTGTGACCTTCGCTGACCCCTGAGCCCTGGGCT 1084
Db 696 SerLeuArgProLeuLeuAlaAlaAlaSerAlaProAlaAlaLeuAla 715
QY 1083 GGCCTGGAGTGTCCCTGTCGGAGGTCA-----GAGTAGCTCCCGACGATACGAA 1030
Db 716 AlAlaAlaAlaSerAlaAlaProAlaArgLugLylArgLysArgLysSerProLylProAla 735
QY 1029 TCACCCGACATCTCGAGACATCCCGCG----- 1003
Db 736 ArgProProLylGylGylGylProArgProProLylThrLylSylSerGylAlaAlaAla 755
QY 1002 -----CAAGTCTCTGTGTCAGACCTAGCTGAGAGAACTCCACGATGG 958
Db 756 ProGylSerAlaAlaAlaAlaProLeuProAlaProAlaProProSerThrPro----- 773
QY 957 ATTGTCCCGACGAGCCCGACCGCTGCTGTCATGTCAGACGAGACGAGGAGCT 898
Db 774 -----ProGylProGylProAlaProAlaGlnProAlaAlaProAlaAlaAlaAla 791
QY 897 TTAGCTACATTCACACTATTGAACAGAGTGGGGCTGCCACGG----- 850
Db 792 GlnAlaArg-----ProArgProValAlaAlaAla 800
QY 849 ---AACCGTCCAGAGGAGGAGGAGCCCGACGTCGAGGAGAGACTAGCTTGCTTAC 793
Db 801 SerArgArgProAlaGlnGylProAlaArgProLeu----- 811
QY 792 AGCCACAGGGTTGAGAA-----GGCAGTCAATACCCCATGATTTCTGGA 748
Db 812 -----GlyLylTrpArgArgGlnProProGylProSerThr 824

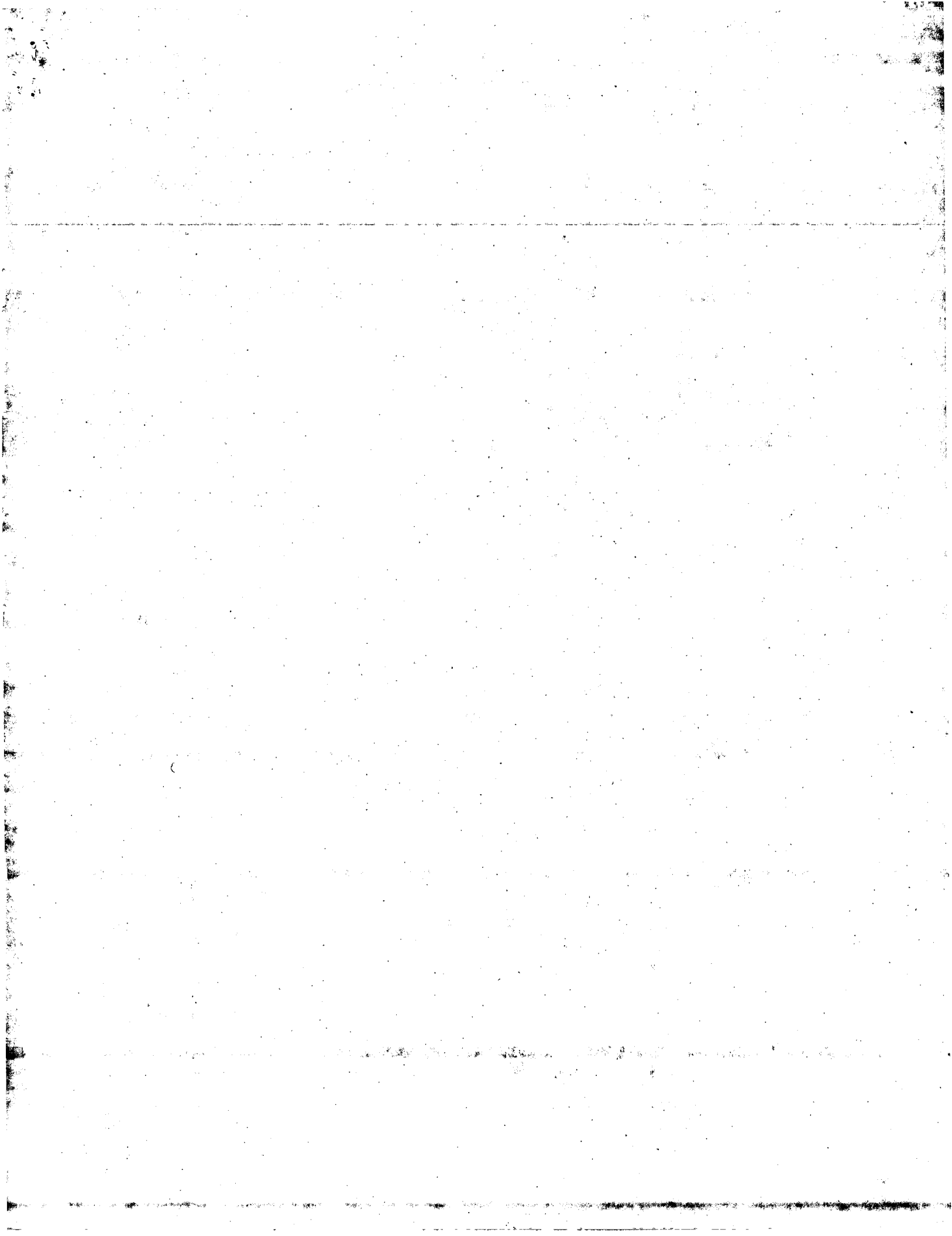
```

```

QY 747 CTAGAGCAGCAGTCCATAGTAGCTGTAACTCACTTTAATTGACATTAT 688
Db 825 -----AlaAlaProAlaAlaAlaAlaLeuGlnAla-----Tyr 835
QY 687 TGCAGCTTTGTCTGCTGACATCTGTCTGTGGAGTGTAGCTTTGTCTGACGGGCTTC 628
Db 836 Cys----- 836
QY 627 CCAGGTGACATGACGTGCTCCAGTGCAGTCACTCTCCACAGGACGAGCCGCTGCT 568
Db 837 -----SerProArgAla 840
QY 567 GAGGATAGTACGTGTGTCAGCCGAGCTTCAGATAGTACGACGAGGCTTCACATATG 508
Db 841 Val---AlaGlnLeuThrAspHisProLeuPheProValProThrArgProAlaLeuMet 859
QY 507 TGCTCTTTGGATGAGGAGCACCTTGCTCCATCTGAGCTTTGAAAAAGACAGCTGCC 448
Db 860 -----PheAspProArgAlaLeuAlaSerIle-----AlaAla 870
QY 447 GACCTGGGGGCGAGAGTGAAGGTAAAGGGCTGATGTTCTTTCTGAGGACAGA 388
Db 871 ArgCysAlaGlyProAlaProAlaAlaGln-AlaAlaCys---GlyGlyGlyAspAspHis 889
QY 387 CCATTCTCCCTAATTGACTGCTTGAGACAGATTAACCTCTCTCATCACTCAGCA 328
Db 889 AspAsnPro-----His 893
QY 327 CCCCTCTGACTTCGATGAGGGGATCACTATCTATCACTCCAGAGAGTCTGCATC 288
Db 893 SPro-----HisGlyAlaAlaGlyGlyArg-----LeuPheGly 904
QY 267 TCCCTTGAAGGGCTTTCCTCCACTGCTCACCCTGTCAGAGTGTGCTGAGACAGA 208
Db 904 yProLeuAlaGlnAlaSerGlyPro-----LeuArgArgMetAlaAlaTrpMetArgGly 921
QY 207 AAGGCCGACACTGGGATGCTCTGCACAGACATCCACAG----- 168
Db 921 nileProAsp-----ProGlnAspValArgValValValle 933
QY 167 -----CTCAGCCAGGTTGAACACTGAAGAGGACAGCTCCGTTGGTGTAGATGCC 115
Db 933 uTyrSerProLeuProGylGlyAspLeuAlaGlyGlyAlaSerGlyGlyProProGyl 953
QY 114 CAGCAGCGAGTGTCTTCGCGGAGTGTCTGTGGATACACTGTACTGCTGACTG 55
Db 953 uTrpSerAlaGlnArgGlyGlyLeuSerCysLeuLeuAlaAlaLeuAlaAsnArgLeuGly 973
QY 54 CTGCTTGCCGTGAGTTCTGACAGATACGCTCCCG-----TCCGATACAGAGTCA 6
Db 973 s-----GlyProAspThrAlaAlaTrpAlaGlyAsnTrpThrGlyAla 987

```

Search completed: June 24, 2003, 14:00:35
 Job time : 50 secs



GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on: June 24, 2003, 13:52:37 ; Search time 51.5 Seconds
(without alignments)
5870.482 Million cell updates/sec

Title: US-09-989-919-15

Perfect score: 2527

Sequence: 1 ggcgtgcacccctaccgga.....aaaaaaaaaaagcggtc 1397

Scoring table: BLOSUM62
Xgapop 10.0, Xgapext 0.5
Ygapop 10.0, Ygapext 0.5
Delop 6.0, Delext 7.0

Searched: 417779 seqs, 108206813 residues

Total number of hits satisfying chosen parameters: 835558

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:
-MODEL=frame+ n2p model -DEV=xld
-Q=/cgn2_1/USFTO.spool/US09989919/runat.24062003.102048.19242/app.query.fasta.1.1543
-DB=published Applications AA -QFMT=fastcan -SUFFIX=rapb -MINMATCH=0.1
-LOOPEXT=0 -UNITS=bits -START=1 -END=-1 -MATRIX=biosum62
-TRANS=human40.cdi -LIST=45 -DOCALLIGN=200 -THR SCORE=pct -THR MAX=100
-THR MIN=0 -ALIGN=1 -MODE=LOCAL -OUTFMT=pco -NORM=ext -HEAPSIZE=500 -MINLEN=0
-MAXLEN=200000000 -USER=US09989919 @CGN 1.1.17 @runat.24062003.102048.19242
-NCPU=6 -ICPU=3 -NO MMAP -LARGEQUERY -NRG_SCORES=0 -WAIT -DSRBLOCK=100
-LONGLOG -DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5
-FGAPOP=6 -FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Published Applications AA:

1: /cgn2_6/prodata/2/pubppa/US08_NEW_PUB.pap:*
2: /cgn2_6/prodata/2/pubppa/PCT_NEW_PUB.pap:*
3: /cgn2_6/prodata/2/pubppa/US06_NEW_PUB.pap:*
4: /cgn2_6/prodata/2/pubppa/US06_PUBCOMB.pap:*
5: /cgn2_6/prodata/2/pubppa/US07_PUBCOMB.pap:*
6: /cgn2_6/prodata/2/pubppa/US07_PUBCOMB.pap:*
7: /cgn2_6/prodata/2/pubppa/PCTUS_PUBCOMB.pap:*
8: /cgn2_6/prodata/2/pubppa/US08_PUBCOMB.pap:*
9: /cgn2_6/prodata/2/pubppa/US09_NEW_PUB.pap:*
10: /cgn2_6/prodata/2/pubppa/US09_PUBCOMB.pap:*
11: /cgn2_6/prodata/2/pubppa/US10_NEW_PUB.pap:*
12: /cgn2_6/prodata/2/pubppa/US10_PUBCOMB.pap:*
13: /cgn2_6/prodata/2/pubppa/US60_NEW_PUB.pap:*
14: /cgn2_6/prodata/2/pubppa/US60_PUBCOMB.pap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysts of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	939	37.2	175	9	US-09-989-919-84 Sequence 84, Appl
2	283	11.1	52	9	US-09-989-919-83 Sequence 83, Appl
3	134.5	5.3	738	9	US-10-057-487-6 Sequence 6, Appl
C	134.5	5.3	738	10	US-09-978-979-6 Sequence 6, Appl

C	5	134	5.2	4123	9	US-10-213-509-5	Sequence 5, Appl
	6	133.5	5.3	1150	9	US-10-123-155-531	Sequence 531, App
	7	130.5	5.2	1346	9	US-10-123-155-481	Sequence 481, App
C	8	122.5	4.8	647	9	US-10-086-464-2	Sequence 2, Appl
C	9	122.5	4.8	647	9	US-10-086-464-4	Sequence 4, Appl
C	10	122.5	4.8	721	9	US-10-086-464-5	Sequence 5, Appl
C	11	121.5	4.7	522	9	US-09-764-866-1138	Sequence 1138, App
C	12	121.5	4.7	522	9	US-09-764-866-761	Sequence 761, App
C	13	120.5	4.7	1336	9	US-10-176-847-68	Sequence 68, Appl
C	14	118	4.6	897	9	US-10-099-899-1	Sequence 1, Appl
	15	117	4.6	2447	9	US-10-123-155-291	Sequence 291, App
	16	116.5	4.6	2226	9	US-10-123-155-355	Sequence 355, App
C	17	116	4.5	957	9	US-10-025-380-1065	Sequence 1065, App
C	18	116	4.5	957	10	US-09-922-217-1065	Sequence 1065, App
C	19	116	4.5	957	10	US-09-833-263-1065	Sequence 1065, App
C	20	116	4.5	1259	9	US-10-260-715-8	Sequence 8, Appl
	21	115.5	4.6	257	9	US-09-813-153-282	Sequence 282, App
	22	114.5	4.5	2782	9	US-10-123-155-205	Sequence 205, App
C	23	112.5	4.4	376	10	US-09-955-518-20	Sequence 20, Appl
C	24	112	4.4	1311	9	US-10-103-377C-6	Sequence 6, Appl
	25	111.5	4.4	1021	9	US-10-184-644-373	Sequence 373, App
	26	111.5	4.4	1021	9	US-10-184-644-373	Sequence 373, App
C	27	111	4.3	707	9	US-09-919-039-278	Sequence 278, App
	28	111	4.4	1231	9	US-10-123-155-279	Sequence 279, App
C	29	111	4.3	5877	9	US-10-142-515-11	Sequence 11, Appl
C	30	111	4.3	5935	9	US-10-243-155-191	Sequence 191, App
	31	110.5	4.4	1076	9	US-10-184-644-191	Sequence 191, App
	32	110.5	4.4	1076	9	US-10-184-644-191	Sequence 191, App
	33	110.5	4.4	3266	9	US-10-123-155-211	Sequence 211, App
	34	110	4.4	1841	9	US-10-184-644-601	Sequence 601, App
	35	110	4.4	1841	9	US-10-123-155-457	Sequence 457, App
	36	110	4.4	1841	9	US-10-184-644-601	Sequence 601, App
	37	110	4.4	3951	9	US-10-184-644-119	Sequence 119, App
	38	110	4.4	3951	9	US-10-184-644-119	Sequence 119, App
C	39	109.5	4.3	1298	9	US-09-825-288A-2	Sequence 2, Appl
	40	109	4.3	1399	9	US-10-184-644-11	Sequence 11, Appl
	41	109	4.3	1399	9	US-10-184-644-11	Sequence 11, Appl
	42	108.5	4.3	1113	9	US-10-123-155-109	Sequence 109, App
C	43	108.5	4.3	1203	9	US-10-067-457-3	Sequence 3, Appl
	44	108.5	4.3	2185	9	US-10-123-155-437	Sequence 437, App
	45	108.5	4.3	2290	9	US-10-123-155-267	Sequence 267, App

ALIGNMENTS

RESULT 1
US-09-989-919-84
; Sequence 84, Application US/09989919
; Patent No. US20020164344A1
; GENERAL INFORMATION:
; APPLICANT: Macina, Roberto
; APPLICANT: Recidom, Herve
; APPLICANT: Pluta, Jason
; APPLICANT: Ghosh, Malavika
; APPLICANT: Sun, Yongming
; APPLICANT: Liu, Chenghua
; TITLE OF INVENTION: Compositions and Methods Relating to Colon Specific Genes and Proteins
; FILE REFERENCE: DEX-0289
; CURRENT APPLICATION NUMBER: US/09/989,919
; PRIOR FILING DATE: 2001-11-21
; PRIOR APPLICATION NUMBER: 60/252,505
; NUMBER OF SEQ ID NOS: 124
; SOFTWARE: Patent version 3.1
; SEQ ID NO 84
; LENGTH: 175
; TYPE: PRT
; ORGANISM: Homo sapien
US-09-989-919-84
Alignment Scores: 2,096-69
Pred. No.: 939.00
Length: 175
Matches: 175

Alignment Scores:	
Pred. No.:	0.012
Score:	134.150
Percent Similarity:	32.09%
Best Local Similarity:	24.11%
Query Match:	5.25%
DB:	10
Gaps:	1
Length:	738
Matches:	136
Conservative:	45
Mismatches:	170
Indels:	174
Gaps:	214

[illegible]

QY	827	GGCCCCAGTGGAGAGACTAGCCCTGAG-	-----	CTTGCTTAACGCCACAGGGT	-----	789	
		:::		:::			
Db	152	euglyleuYrTrrYrserAlaenGluGInCysArgValAlaPheGlyProGlySalAValA	172				
QY	780	-----TGGAGAGCAGCTCATPACCCCATGATTCCTGGACTAGA-	-----			742	
Db	172	IaCysThrPheAlaIaArgGluHisIaLeuValSerLeuProAlaValAlaTrpSerPheU*	192				
QY	741	-----GGCAGCAGTCCCATAGAGGCTGGTAACTACACATTTTAACTTTGGCATT	698				
Db	192	**GlyProSerAlaSerProSerSerArgProPheU-	-----	ArgAlaTrp-Ile	208		
QY	689	ATTGACGTTTGTCTGTGTCACATCTGTCTGTGGAGTGTGAATTCCTTGTCGACGGGCT	638				
Db	209	CysAlaArgProSerProAlaThrGlnThrArgTrp-	-----			220	
QY	629	TCCAGGTGACATGAGCTGCTCCAGTCCAGTACATCTCCCTCCACAGGACCCACTG	570				
Db	221	-----ThrValAlaAlaAlaAlaAlaAlaSerPheLeuSerTrpMetGlyGlnAsn	237				
QY	569	CTG-	-----	AGCATPAGCAC-	-----	TGTCAGCCGAGCCCTCA	537
		:::		:::			
Db	238	ValAlaIaTrpArgSerGlyAlaProArgValAlaAlaAlaPheTrpTrpSer***ProPro	257				
QY	536	GATAGGTCAGCCAGAGCGCTTCACATATGCGTCTTGTGGATCAGGACCACTTGGCT	477				
Db	258	**GlnGlnCysMetGly-----AlaGlyLeuAlaGlyAlaProGluValIleAla	274				
QY	476	CCATCCACGCTTGGAAAAGACAGCAGCTGCCAGCTGGGGGGGAGAGGTGACAGGTAAGG	417				
Db	275	ProIaPheProAlaGluGlnValaIaTrpSerPro-GlyGlyGly-	-----	SerAlaThrTh	291		
QY	416	GCTGATGTTCTCTTTCTGAGGAGACCAATCTC-	-----			380	
Db	291	rProAspLeuProLeuGlyGlyValaHisIaValLeuValLeuThrSerArgProArgCysAl	311				
QY	379	-----GCCAATTT	372				
Db	311	AlThrLeuAlaGProAlaIaArgAlaProSerTrpSerSerCysArgAsnSerAlaProGlyPr	331				
QY	371	GACTGCGCTTGAAGACAGCATATCTCTCTC-	-----			341	
Db	331	OTThrAlaSerArgCysAlaProProLeuAlaAlaProProSerThrThrGlyValLeuLe	351				
QY	340	-----CATCACCTCAGACACCCCTCGACTTCCGAT-	-----			310	
Db	351	uTrpHis-ThrAlaGlySerLeuMetCysAlaAspThrCysAlaGlyProLeuAlaArgA	371				
QY	309	-----GGGGGATCAGTATCTATCAACTCCAGAGAGGCTGCC-	-----	ATGTC	264		
Db	371	IaSerSer***SerValGlnThrAlaSerSerMetGlyProGlyValCysGlnValAlaIaP	391				
QY	263	TGGAGAGGCTTCTCTCC-	-----	ACTGAGCTCAGTCTCCAGG	228		
Db	391	roGlyArgThrGlyPro***AlaCysValCysArgAlaAlaIaGly-HisLeuAlaVal	410				
QY	227	TGGTGTGTTGTGACACACAAAGCGCCGACCTGG-	-----			192	
Db	411	MetValGlyTrpThrProSerArgTrGlyIaThrGlyAlaArgCysValValGlyThrThr	430				
QY	191	-----CATGGCTCT	-----			183	
Db	431	AlaArgAlaAlaHisIGlyArgAlaLeuSerGlnLeuAlaGlnArgGluAsnMetSerArg	450				
QY	182	-----CACAGACATCCACAGCTCAGCCAGGTTGAACATG-	-----	AAAGGAGCAGCTCC	132		
Db	451	Phe***GlnLeuProProThr***ProValSerThrIleuProThrThrGlyLeuSerSer	470				
QY	131	CGTGTGTGTGATGAGCAGCAGCGAGTACT-	-----			100	
Db	471	HisThrTrpArg***GlySerGlnGlyAlaIaMetSerTrpLeuGlyArg***AlaSerPro	490				

QY 99 -----TCCTGGGGAGTGGTCTGTCTGGAGTACACTGGTACTCG 61
 Db 491 LeuThProProThrProProSerTyrPargMetValValSerThrGluTrpProSer 510
 RESULT 5
 US-10-213-509-5
 ; Sequence 5, Application US/10213509
 ; Publication No. US20030054485A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Weiss, Joseph
 ; APPLICANT: Scott, Matthew
 ; TITLE OF INVENTION: JELLY BELLY GENES AND THEIR USES
 ; FILE REFERENCE: STAN-232
 ; CURRENT APPLICATION NUMBER: US/10/213,509
 ; CURRENT FILING DATE: 2002-08-06
 ; PRIOR APPLICATION NUMBER: 60/311,720
 ; PRIOR FILING DATE: 2001-08-09
 ; NUMBER OF SEQ ID NOS: 5
 ; SOFTWARE: FastSeq for Windows Version 4.0
 ; SEQ ID NO 5
 ; LENGTH: 4123
 ; TYPE: PRT
 ; ORGANISM: H. sapiens
 ; US-10-213-509-5
 Alignment Scores:
 Pred. No.: 0.0197 Length: 4123
 Score: 134.00 Matches: 110
 Percent Similarity: 33.41% Conservative: 36
 Best Local Similarity: 25.17% Mismatches: 134
 Query Match: 5.23% Indels: 157
 Gaps: 26
 US-09-989-919-15 (1-1397) x US-10-213-509-5 (1-4123)
 QY 1249 GGGGAAAGCTGCTTTGGCTGCTTGGAGCTTTCATGAAACCAATGAGCAAGAG 1190
 Db 1799 GtGtAlaAlaLeuProSerGlySerLeuValLeuSerLeu-----AspArg 1813
 QY 1189 CCCCCGCCCCAGCCCGCCAGTCTGCTGCAACCAACCACTGCTCTGACCCCACT 1130
 Db 1814 ProAlaAlaHisProProProProSer-GlySerAspCysTrpProSerLeuSerGly 1833
 QY 1129 CTGGCTCATGTAGTGTGACCTCTGCTGCTGACCTGACCTGCTGCTGCTGCTG 1076
 Db 1833 u-TripLeuValLeu---LeuValThrLeuGlyGlnValProGlyProLeuTrpLeuProG 1852
 QY 1075 AGTGTCTCC-----CTGTGGGAGGTCAGAGATAGCTCCCAAGTACAGAAATCACCAC 1022
 Db 1852 LnhIsProValValLeuProGluLeuGlnInProProProLeuArgProArgSerProV 1872
 QY 1021 ATGCTGAGATCCCGCCCAAGTCTCTCCAGACT-----AAG 980
 Db 1872 alProTrp-HisProProGlnGlyTrpGlnThrGluProCysGlnGlyGln 1891
 QY 979 CTGAGAGAACACTCCAGATGATGTCACAGACAGCCCAAGCTGCTGCTGAGATG 920
 Db 1892 GlyGlnValHisArgValGlyGlnArgTrpHisGlyGlyPro-----CysArgVal 1908
 QY 919 TGCAGACAGCAGCAGAGGGGCTTATGATCTCACTATATGAAACAGATGATGGG 860
 Db 1909 CysGln-CysLeuHisAsnLeuThrAlaHisCysSerProTy----- 1922
 QY 859 CTGCCCAAGGAGACCTGCCAGGAGGGGACACACCCCACTGGGGAGACTAGCTT 800
 Db 1923 -CysPro-----Le 1925
 QY 799 TGCCTACAGCCACAGGCTTGAAGAGCAGCTACATACCCCATGATTCCTGAGTACGAGC 740
 Db 1925 uGlySerCysProGlnGlyTrp-----ValLeuValGlnGlyThrGlyGln 1940
 QY 739 CAGCAGTCCCATAGTACTGCTGTAACATCAATTATTTGACATTTGACAGCTT 680
 |||

Db 1940 user----- 1941
 QY 679 TGTCTGTGACATCTGTCTGTCTGGAGTGTAGCTTTGTCTGACAGGGTCCAGATGA 620
 Db 1942 -----CysCysHisCys-----AlaLeuProGlyGln 1950
 QY 619 CATGAGTCTGTCAGTGCAGTCACTGCTCT-----CCACAGGCAAGCCCACTGTC 569
 Db 1950 uAsnGlnThrValGlnProMetAlaThrProAlaAlaProAlaProSerProGln-1 1970
 QY 568 TGAGG-----ATAGTCACTGTGTGACCGGAGCCCTCAATAGTGTACAGCAGAGGCTTC 515
 Db 1970 LeArgHeProLeuAlaThrTyThrLeuProProSerGlyGlySerCysArgProLeu 1990
 QY 514 AC-----ATATGT-----G 506
 Db 1990 eSerProThrProAlaCysLeuSerLeuHisProAspProCysTyTySerProLeuG 2010
 QY 505 GTCTTGTGGATCAGGAGC-----CACTTGCCTCCATCCAGCTTGTGAAAGAGCC 455
 Db 2010 TyLeuAlaGlyLeuAlaGlnGlySerLeuHisAlaSerSerGlnGlnGlnHisProT 2030
 QY 454 AG-----CTGCCGACCTGGGGGGCAGAGGAGAGAGTAAACAGGCTG 413
 Db 2030 hrcGlnAlaAlaLeuLeuGlyAlaProThrGlnGlyProSerProGlnGlyTrpHisAlaG 2050
 QY 412 GA-----TGTTCCTCTTCTGAGAG-----CAAGACCACTTCTCCATATTGACTGCTG 362
 Db 2050 TyGlyAspAlaTyAlaAlaTyTrpHisThrArgProHisTyThrLeuGlnLeuLeuLeuG 2070
 QY 361 AA-----GACACGTAACT 347
 Db 2070 LnhProArgAsnLeuThrGlyThrLeuValProGluThrGlySerSerAsnAlaTyAlaS 2090
 QY 346 CTCTCTCATCACTCAGACACCCCTGTGACTTCCATGGGGATCACTATATCAACT 287
 Db 2090 eSerPheSerLeuGlnPheSer-----SerAsnGlyLeuHisTrp---HisAsp 2106
 QY 286 CCAGAGAGGTCCTGACCATCTCCCTTGAAAGGCTTCCCACTGCTGCTGCTGCTG 221
 Db 2106 yTrpGAspLeuLeuPro-----GlyLeuLeuProLeuProLysValSerProA 2122
 QY 230 -----AGTGTCT-----GGTGTGTAC 212
 Db 2122 LaglnGlyArgTrpGlyGlnGlnInProThMetProPheCysGlyPheHisSerLeuCy 2142
 QY 211 CACAAAGCCCGCACT-----GGCAGTGGCTTCAAC 180
 Db 2142 roGlnGlyProSerSerValProGlnGlyHisGlyLeuHis 2155
 RESULT 6
 US-10-123-155-531
 ; Sequence 531, Application US/10123155
 ; Publication No. US20030068794A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Baker, Kevin P.
 ; APPLICANT: Beresini, Maureen
 ; APPLICANT: Deforge, Laura
 ; APPLICANT: Desnoyers, Luc
 ; APPLICANT: Filvaroff, Ellen
 ; APPLICANT: Gao, Wei-Qiang
 ; APPLICANT: Gerritsen, Mary E.
 ; APPLICANT: Goddard, Audrey
 ; APPLICANT: Godowski, Paul J.
 ; APPLICANT: Gurney, Austin L.
 ; APPLICANT: Sherwood, Steven
 ; APPLICANT: Smith, Victoria
 ; APPLICANT: Stewart, Timothy A.
 ; APPLICANT: Tumas, Daniel
 ; APPLICANT: Watanabe, Colin K
 ; APPLICANT: Wood, William
 ; APPLICANT: Zhang, Zemin
 ; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC

QY 123 ACCACGAGGAGCTGCTCTTCAGTGTTCACCTG---GCTGAGCTGTGATGTC 178
 |||||
 Db 458 ThrThrThrGlyCysAlaThr-----CysAlaThrCysAlaThrAlaGlyThrThrGly 475
 QY 179 TGTGAGAGCATGCTCCAGTGTGGGCTTTGTGTACCAACAGACCTCTGACAGGT 238
 |||||
 Db 476 CysAlaAlaThrThrGlyCysAlaCysThrAlaCysThrGlyAlaThrThrThrAla 495
 QY 239 GAGCCAGTGGAGAGAGCCCTTCAGAGGAGATGGCAGGAGCTCTGAGGTGATAGAT 298
 |||||
 Db 496 ThrCysAlaGly----- 499
 QY 299 AGTATGCCCATCCGAGAGTCAAGGGGGTCTGAGGTGATGAGAGAGATACGTGT 358
 |||||
 Db 500 -----GlyGlyAla-----ThrCys 504
 QY 359 CTTCAGAGC---AGTCAATTAAGGAGATGTCTTCTCCAGAAAGAGAAACATTCAG 415
 |||||
 Db 505 ThrGlyGlyCysAlaAlaCysGlyThrAlaGlyAlaAlaGly----- 518
 QY 416 CCTGTACTCTCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 475
 |||||
 Db 519 -----AlaAlaAlaGlyAla----- 523
 QY 476 GAGCCAGTGTCTCTGATCCCAACAGACATATGTGAAGGCTCTGCTGATCTAT 535
 |||||
 Db 523 ----- 523
 QY 536 CTGAGGCTGCTGCTGACAGCTGATCTCTCAAGCTGCTGCTGCTGCTGCTGCTGCT 586
 |||||
 Db 524 -----AlaCysAlaAlaAlaGlyAlaAlaGlyCysAlaThr 535
 QY 587 -----GAGGAGAGTACT-----TGCACTGGC----- 608
 |||||
 Db 536 CysThrGlyAlaAlaGlyThrGlyGlyAla-ThrGlyAlaCysGlyCysThrGlyAlaAl 555
 QY 609 -----AGCAGCTGATGTCACCTGGAAACCC 634
 |||||
 Db 555 AGlyAlaThrAlaAlaGlyThrGlyThrGlyAlaAlaAlaAlaCysAlaThrGlyAlaThr 575
 QY 635 TGCAGCAAGCTTACATCCCAAGACAGATGTGACCAAGCA----- 680
 |||||
 Db 575 rCysAlaCysAlaAlaThrThrGlyAlaAlaAlaAlaThrGlyCysAlaThrCysCys 595
 QY 681 -----ACGTGCAATATGCAAACTTAAATGTGAGTTACAGCTTACGCTATGGA 733
 |||||
 Db 595 sCysCysThrCysThrGlyAlaThrCys----- 604
 QY 734 CTGCTGAGCTCTAGTCCAGGAATCATGGGGGTATGACTGCTCTCCAACTGTGGGCTG 793
 |||||
 Db 605 -----CysCysCysThrGlyAlaCys 612
 QY 794 TAAGCAAGCTGAGCTAGTCTCCCACTGGGGGCTGTGCTCTCCCTGGAGCGTTCCGT 853
 |||||
 Db 612 sAlaThrGlyAlaAlaGlyGlyGlyGlyCysAlaThrAlaThrAlaAlaThr 632
 QY 854 GGGGAGCCCATCATCTGTTCATATGAGATGTGAGATGTACTAAAGCCCTGCTC----- 906
 |||||
 Db 632 rGly-AlaThrGlyCysCysThr-----ThrCysAlaThrGlyAlaCysAlaGlyAla 649
 QY 907 -----TGCTGCTGCTGCAAT 922
 |||||
 Db 649 laGlyGlyAlaThrGlyAlaGlyAlaGlyCysThrCysAlaCysCysCysCysThrC 669
 QY 923 GCCACA-----GCAGCGGTGGGGGCTGCTGGGGAACAATCCATCTGAGAGTTC 973
 |||||
 Db 669 ys-ThrCysThrGlyAlaAlaGlyGlyCysThrGlyThr----- 682
 QY 974 TCTCAGTTAGTCTGAGACAGAGACTTGGCGGGGATGCTCCAGAGATGCGGTATCT 1033
 |||||
 Db 683 -----ThrGlyThrThr 686

QY 1034 GTAAGTGGGAGGCTATCTTCTGACTCCGAGAGGGGAGACTCCAGCCAGCGAGG 1093
 |||||
 Db 687 Cys-----ThrGlyCysThrCysCysThrCysAla 697
 QY 1094 TCAGGGGAGAGGTGACACTTCAGATGAGCCAAACCTGGGTACAGGAGAGGTGTG 1153
 |||||
 Db 698 AlaGlyAlaAlaAlaThrThrAlaAlaAlaCysAlaThrThrThrGlyThrThrCys 717
 QY 1154 TTGAGCCAGAGACTGGGCGGGGGTGGG-----CCGGGCTTCTGCTCATTTGC 1207
 |||||
 Db 718 ThrGlyThrGlyThrGlyAlaCysThrGlyCysThrGlyAlaGlyCysAlaThrCys 736
 RESULT 8
 US-10-086-464-2
 ; Sequence 2, Application US/10086464
 ; Publication No. US20020199218A1
 GENERAL INFORMATION:
 ; APPLICANT: GORING, Daphne R. et al.
 ; TITLE OF INVENTION: PROLINE-RICH EXTENSIN-LIKE RECEPTOR KINASES
 ; FILE REFERENCE: P 25,762-A USA
 ; CURRENT APPLICATION NUMBER: US/10/086,464
 ; PRIOR FILING DATE: 2002-02-28
 ; PRIOR APPLICATION NUMBER: US 10/069,304
 ; PRIOR FILING DATE: 2002-02-19
 ; PRIOR APPLICATION NUMBER: PCT/CA00/00966
 ; PRIOR FILING DATE: 2000-08-18
 ; PRIOR APPLICATION NUMBER: US 60/149,466
 ; PRIOR FILING DATE: 1999-08-19
 ; PRIOR APPLICATION NUMBER: US 60/159,122
 ; PRIOR FILING DATE: 1999-10-13
 ; NUMBER OF SEQ ID NOS: 27
 ; SOFTWARE: Patentin Ver. 2.1
 ; SEQ ID NO 2
 ; LENGTH: 647
 ; TYPE: PRT
 ; ORGANISM: Brassica napus
 US-10-086-464-2
 Alignment Scores:
 Pred. No.: 0.107 Length: 647
 Score: 122.50 Matches: 91
 Percent Similarity: 31.28% Conservative: 36
 Best Local Similarity: 22.41% Mismatches: 129
 Query Match: 4.79% Indels: 150
 DB: 9 Gaps: 19
 US-09-989-919-15 (1-1397) x US-10-086-464-2 (1-647)
 QY 1188 CCCGCCCCAGCCCGCCGAGGTCTGCTCAACACACTGCTCCCTGACCCGAGTC 1129
 |||||
 Db 7 ProGlyThrGlySerProProSerProProSerProProSerProProSerProPro 25
 QY 1128 TTGGCTCATGCTGAGGTGTGACCC-----TTGCCCCCTGACCCCTGGGCTGCTGGA 1075
 |||||
 Db 26 ProAlaSerAlaProProProProProProProProProProProProProProPro 45
 QY 1074 GTGTCC-----CCTTGGGAGGTGAGATAGCTCTCCAGAGTACAGATACACCA 1021
 |||||
 Db 46 ThrSerProProProSerSerSerSerThrProSerAlaProProProProProPro 65
 QY 1020 TCCTGAGCATCCCC-----CGCAAGTCTCTGTCCAGACTTAAGC 979
 |||||
 Db 66 Pro---SerThrProGlySerProProProProleuProGlnProSerProProAlaProThr 84
 QY 978 TGAGAGAACTTCACAGATGATGTTC-----CCAGCGAGCCGCCACCGCTGTGTG 925
 |||||
 Db 85 ThrProGlySerProProAlaProValThrProProThrArgAsnProProProSerVal 104
 QY 924 GCATGTGACAGCAGACAGCAGGAGGCTTAACTCACTTCTCACTATTGAACACAGTGA 865
 |||||
 Db 105 Pro----- 105
 QY 864 TGGGCTGCCACGAAACCGTCCAGAGAGGGGACAGACCCCGAGTGGAGAGTACGCT 805

```

106 ---GlyProProSerSerProSerArgGlyGlySerProArg-----Pro 120
QY 804 GAGCTTGCTTACAGCCACAGGGTTGAGAGGAGCATACCCCATGATTCCTGACTA 745
Db 121 ProSerSerProSerProPro-----SerProSerSerArgGlyLeu 134
QY 744 GAGAGCCAGTCCCATAGTAGGCTGTGTAACATCTTTAACTTTGGCATTTATG 685
Db 135 SerThrGlyValValAlaGlyLeuAlaIleGlyGly-----ValAla 148
QY 684 ACCTTTGCTGCTGCTCATCTGTCTGTGGATGTTAGCTTTGTCTGAGGGGTTCCCA 625
Db 149 LeuLeuValIleValIleThrLeuIleCysLeu-----LeuGlySylSylSylSylArgArg 165
QY 624 GGTGACATGCAAGTGTCTGCCAGTGCAGAGTCACTCCCTCCACAGGCAAGCCAGCTGCTGAG 565
Db 166 ArgAspGlnGlnAspAlaTyrTyrValProProProProProProGlyProLysAla--- 184
QY 564 GATAGTCAGCTGTGTCAGCCGAGCCCTCAGATAGTACGACGAGGCTTTCATATATGTG 505
Db 184 ----- 184
QY 504 TCTTGTGGATCAGGAGCAGCTTGCTCCATCCAGTGAAGAAAGACAGCTGCCGAC 445
Db 185 -----GlyGlyProTyrGlyGlyGlnGlnGln----- 194
QY 444 CTGGGGGAGAGGTGAGAGTAAAGGGCTGATGTTCTCTTCTGAGGCAAGACCA 385
Db 195 -----TyrArgGlnGlnArg 199
QY 384 TTTCCCTATTGACTGCTTGAAGACAGTATACCTCTCTCATCAGCTCAGACCC 325
Db 199 nAla----- 200
QY 324 CCTCTGACTTCCGATGGGAGTCACTATCTCACTCCAGAGAGTCCCTGCCATCTCC 265
Db 201 -----ThrProProSerSerAspHisValIleThrSer-----LeuProProPro 214
QY 264 CTGGAAGAGGCTTCTCCCATCTGCTCACTGCTCAGGTGCTGTGTTGTGACCAAG 205
Db 214 oProLysAlaProSerPro-----ProArgI 223
QY 204 GCCCGACACTGGGAGTGGCTCTCAGACATCCACAGCCTCAGCCAGGTGAACACTGAA 145
Db 223 nProProPro-----ProProProProProPhenylSerSerSer 236
QY 144 AGGAGCAGCTCCCGTGTGTGTAGATGAGCAGCAGCGGTAGTCTTCTGGGGAGTGTG 85
Db 236 rGlyGlySerAspLysSerAspArgProValLeuProProProSerProGly-----Le 254
QY 84 CTGCTGGGATPACCTGTAAGTGTGTAAGTCTGCTGCTGCGGTGAGTTCTGACATATG 25
Db 254 uValLeuGlyPheSerLysSerThrPhe-----ThrTyrGlnGlnLeuAl 269
QY 24 CCGCTCCGATACAGT 9
Db 269 aArgAlaThrAsnGly 274

RESULT 9
US-10-086-464-4
; Sequence 4, Application US/10086464
; Publication No. US20020199218A1
; GENERAL INFORMATION:
; APPLICANT: GORING, Daphne R. et al.
; TITLE OF INVENTION: PROLINE-RICH EXTENSIN-LIKE RECEPTOR KINASES
; FILE REFERENCE: P 25,762/A USA
; CURRENT APPLICATION NUMBER: US/10/086,464
; CURRENT FILING DATE: 2002-02-28
; PRIOR APPLICATION NUMBER: US 10/069,304
; PRIOR FILING DATE: 2002-02-19
; PRIOR APPLICATION NUMBER: PCT/CA00/00966
; PRIOR FILING DATE: 2000-08-18

```

```

; PRIOR APPLICATION NUMBER: US 60/149,466
; PRIOR FILING DATE: 1999-08-19
; PRIOR APPLICATION NUMBER: US 60/159,122
; PRIOR FILING DATE: 1999-10-13
; NUMBER OF SEQ ID NOS: 27
; SOFTWARE: Patent In Ver. 2.1
; SEQ ID NO 4
; LENGTH: 647
; TYPE: PRT
; ORGANISM: Brassica napus
; US-10-086-464-4

Alignment Scores:
Pred. No.: 0.107 Length: 647
Score: 122.50 Matches: 91
Percent Similarity: 31.28 Conservative: 36
Best Local Similarity: 22.41% Mismatches: 129
Query Match: 4.79% Indels: 150
DB: 9 Gaps: 19

US-09-989-919-15 (1-1397) x US-10-086-464-4 (1-647)
QY 1188 CCGGCCCCACCCCGCCAGTCTGTGCTCAACACACCTGCTCCAGCCGATC 1129
Db 7 ProGlyThrGlySerProProSerProProSerProSerProSerProProSerPro 25
QY 1128 TTGGCTCATGCTGAGGTGCAAC-----TTGGCCCTGACCCCTGGGCTGCGTGGGA 1075
Db 26 ProAlaSerAlaProProProProProProProProProProProProProProPro 45
QY 1074 GTGTCC-----CCTGTGGGAGGTGAGATAGCTCCCTCCAGTACAGATACACCCACA 1021
Db 46 ThrSerProProProSerSerSerSerSerThrProSerAlaProProProSerProPro 65
QY 1020 TCTTGAGCATCC-----CCCAAGTCTCTGTCCAGCTAAGC 979
Db 66 Pro-----SerThrProGlySerProProProLeuProGlnProSerProProAlaPro 84
QY 978 TGAGAGAACATCCACGATGATGTC-----CCAGCAGACCCCGCCGCTGCTGTG 925
Db 85 ThrProGlySerProProAlaProValIleProProThrArgAsnProProProSerVal 104
QY 924 GCATGTGACAGACAGCAGCAGGAGGCTTGTAGTACTTCTCAGACTATTGAACAGTGA 865
Db 105 Pro----- 105
QY 864 TGGGCTGCCACGAAACGCTCCAGAGGAGGAGGAGCAGCCCGCAGTGGGAGACTAGCT 805
Db 106 ---GlyProProSerSerProSerArgGlyGlySerProArg-----Pro 120
QY 804 GAGCTTGCTTACAGCCACAGGGTTGAGAGGAGCATACCCCATGATTCCTGACTA 745
Db 121 ProSerSerProSerProPro-----SerProSerSerArgGlyLeu 134
QY 744 GAGAGCCAGTCCCATAGTAGGCTGTGTAACATCTTTAACTTTGGCATTTATG 685
Db 135 SerThrGlyValValAlaGlyLeuAlaIleGlyGly-----ValAla 148
QY 684 ACCTTTGCTGCTGCTCATCTGTCTGTGGATGTTAGCTTTGTCTGAGGGGTTCCCA 625
Db 149 LeuLeuValIleValIleThrLeuIleCysLeu-----LeuGlySylSylSylSylArgArg 165
QY 624 GGTGACATGCAAGTGTCTGCCAGTGCAGAGTCACTCCCTCCACAGGCAAGCCAGCTGCTGAG 565
Db 166 ArgAspGlnGlnAspAlaTyrTyrValProProProProProProGlyProLysAla--- 184
QY 564 GATAGTCAGCTGTGTCAGCCGAGCCCTCAGATAGTACGACGAGGCTTTCATATATGTG 505
Db 184 ----- 184
QY 504 TCTTGTGGATCAGGAGCAGCTTGCTCCATCCAGTGAAGAAAGACAGCTGCCGAC 445
Db 185 -----GlyGlyProTyrGlyGlyGlnGlnGln----- 194

```

QY 444 CTGGGGGAGAGAGTGAAGTAACAGGGCTGATGTTCTCTTTCTGAGGACAGCA 385
 Db 195 -----TriprarglnGln 199
 QY 384 TTCTCCCTAATTGACTGCTTGAAGACAGTATACCTCTCTCTACCTACGACACC 325
 Db 199 naLa ----- 200
 QY 324 CCTCTACTTCCGATGGGGATCACTATCTATCACTCCAGAGAGTCCGCTCATCTCC 265
 Db 201 -----ThrProProSerAspHisValValInrSer-----LeuProProPr 214
 QY 264 CTGGAAGGGCTTCTCCCACTGCTCACTGCTCCAGGTGCTGTGTGTGACACAAAG 205
 Db 214 oProLysAlaProSerPro-----ProArgG 223
 QY 204 GCGCCGACACTGGGAGCTCTTCACAGACATCCACAGCTCAGCCAGTTGAACACTGAA 145
 Db 223 nProProPro-----ProProProProPhemeSerSerSe 216
 QY 144 AGGAGCAGCTCCCGTGTGTGTAGATGCGCAGCAGCGGTAGCTTCTCCGGGGAGTGTG 85
 Db 236 rGlyGlySerAspTyrSerAspArgProValLeuProProProSerProGly-----Le 254
 QY 84 CTGTCTGGATACACTGCTACTGCTGCTGCTGCTGCGTGGAGTTCTGACAGACTGTC 25
 Db 254 uValLeuGlyPheSerLysSerThrPhe-----ThrTyrGlnGlnLeuAl 269
 QY 24 CCGCTCCGCTGACAGT 9
 Db 269 aArgAlaTrnAsnGly 274

RESULT 10

US-10-086-464-5
 : Sequence 5, Application US/10086464
 : Publication No. US20020199218A1
 : GENERAL INFORMATION:
 : APPLICANT: GORING, Daphne R., et al.
 : TITLE OF INVENTION: PROLINE-RICH EXTENSIN-LIKE RECEPTOR KINASES
 : FILE REFERENCE: P 25,763-A USA
 : CURRENT APPLICATION NUMBER: US/10/086,464
 : CURRENT FILING DATE: 2002-02-28
 : PRIOR APPLICATION NUMBER: US 10/069,304
 : PRIOR FILING DATE: 2002-02-19
 : PRIOR APPLICATION NUMBER: PCT/CA00/00966
 : PRIOR FILING DATE: 2000-08-18
 : PRIOR APPLICATION NUMBER: US 60/149,466
 : PRIOR FILING DATE: 1999-08-19
 : PRIOR APPLICATION NUMBER: US 60/159,122
 : PRIOR FILING DATE: 1999-10-13
 : NUMBER OF SEQ ID NOS: 27
 : SOFTWARE: PatentIn Ver. 2.1
 : SEQ ID NO 5
 : LENGTH: 721
 : TYPE: PR1
 : ORGANISM: Brassica napus
 : US-10-086-464-5

Alignment Scores:

Pred. No.: 0.11 Length: 721
 Score: 122.50 Matches: 91
 Percent Similarity: 31.28% Conservative: 36
 Best Local Similarity: 22.41% Mismatches: 129
 Query Match: 4.79% Indels: 150
 Db: 9 Gaps: 19

US-09-989-919-15 (1-1397) x US-10-086-464-5 (1-721)

QY 1188 CCGGGGCCACACCCCGCCAGGCTCTGCTCAACACACCTGCTCCCGACCCAGTC 1129
 Db 37 ProGlyThrGlySerProProSerProProSerAsnSerThr--ThrThrThrProPro 55

QY 1128 TTGGCTCATGCTGAGGTGTGACCC-----TGTGCCCTGACCCCTGGGCTGGCTGGGA 1075
 Db 56 ProAlaSerAlaProProProThrThrProSerSerProProProSerThrIlePro 75
 QY 1074 GTGTCC-----CCTGTGGGAGTGAAGATAGCTCCCGACAGTACAGATACCCCA 1021
 Db 76 ThrSerProProProSerSerAspSerThrProSerAlaProProProSerProThr 95
 QY 1020 TCTTGAGCATCCCC-----CGCCAGTCTCTGTCCAGCATTAAGC 979
 Db 96 Pro--SerThrProGlySerProProProLeuProGlnProSerProProAlaProThr 114
 QY 978 TGAGAACACTCCACAGATGATGTC-----CCGACGACGCCCCAGCCGCTGCTGTG 925
 Db 115 ThrProGlySerProProAlaProValThrProProThrArgAsnProProSerVal 134
 QY 924 GCATGTGCACGACGACGACGAGGGGCTTAACTACATTCTACACATTAATGAACAGTGA 865
 Db 135 Pro----- 135
 QY 864 TGGGGCTGCCACGGAACCGTCCAGGGAGGGGACAGCCCGAGTGGGAGACTAGCT 805
 Db 136 ---GlyProProSerAsnProSerArgGlnGlySerProArg-----Pro 150
 QY 804 GAGCTTCTTAAAGCCACAGAGGTTGGAGAGGACGTACACCCCATGATTCCTGAGACTA 745
 Db 151 ProSerSerProSerProPro-----SerProSerSerAspGlyLeu 164
 QY 744 GAGCCAGCAGTCCCATGACTGAGCTGTGAACACTCATATTTAATTTGGCATTAATGC 685
 Db 165 SerThrGlyValValValGlyIleAlaIleGlyIly-----ValAla 178
 QY 684 ACCTTGTCTGCTGATCATCTGCTGCTGCTGAGATGTAGCTTTGTGACAGGGTCCCA 625
 Db 179 LeuLeuValIleValThrLeuIleCysLeu-----LeuCysLysLysValArg 195
 QY 624 GGTGACATCAGTGTGCTCCAGTGAAGTCACTCCCTCCACAGGCAACCCAGCTGCTGAG 565
 Db 196 ArgAspGlnGlnAspAlaTyrTyrValProProProProProGlyProLysAla-- 214
 QY 564 GATATGACTGTGTACGCCGAGCCCTCAGATAGTGCACGACAGGCTTACATATGTGG 505
 Db 214 ----- 214
 QY 504 TCTTGTGGATCAGGAGACCACTGTGCTCCATCCAGCTTTGAAGAACACAGCTGCCGAC 445
 Db 215 -----GlyGlyProTyrGlyGlnGlnGln--Gln-- 224
 QY 444 CTGGGGGAGAGTGAAGTGAACAGGCTGATGTTCTCTTCTGAGGACAGCA 385
 Db 225 -----TriprarglnGln 229
 QY 384 TTCTCCCTAATTGACTGCTTGAAGACAGTATACCTCTCTCTACACTTCAGACACC 325
 Db 229 naLa ----- 230
 QY 324 CCTCTACTTCCGATGGGGATCACTATCTATCAACTCCAGAGAGTCTGCTCATCTCC 265
 Db 231 -----ThrProProSerAspHisValValInrSer-----LeuProProPr 244
 QY 264 CTGGAAGGGCTTCTCCCACTGCTCACTGCTCCAGGTGCTGTGTGTGACACAAAG 205
 Db 244 oProLysAlaProSerPro-----ProArgG 253
 QY 204 GCGCCGACACTGGGAGCTCTTCACAGACATCCACAGCTCAGCCAGTTGAACACTGAA 145
 Db 253 nProProPro-----ProProProProPhemeSerSerSe 266
 QY 144 AGGAGCAGCTCCCGTGTGTGTAGATGCGCAGCAGCGGTAGCTTCTCCGGGGAGTGTG 85
 Db 266 rGlyGlySerAspTyrSerAspArgProValLeuProProProSerProGly-----Le 284
 QY 84 CTGTCTGGATACACTGCTACTGCTGCTGCTGCTGAGTTCTGACAGTACTGC 25

Alignment Scores:

Pred. No.: 0.122 Length: 524
 Score: 121.50 Matches: 105
 Percent Similarity: 32.63% Conservative: 35
 Best Local Similarity: 24.48% Mismatches: 128
 Query Match: 4.75% Indels: 161
 DB: 9 Gaps: 27

US-09-989-919-15 (1-1397) x US-09-764-868-761 (1-524)

QY 1191 GGGCCCCGGCCCCCAGCCCC---GCCCGAGTCTGGCTCAACACACACCTGCTCCCTGACC 1135
 DB 78 GlySerGlyProGluProLeuAlaProSerPro-----ValSer 90
 QY 1134 CCAAGTCTGGCTATGCTGAGGTGTGACCTGACCCCTGACCCCTGGGCTGGCGGA 1075
 DB 91 ProThrPhePro-----ProSerSerProSerAspTrpPro---GlnGlu 104
 QY 1074 GTGTCCCTGTCCGAGAGTCAGATAGCTCCCGACAGTACAGATCACCCACATCC--- 1018
 DB 105 ArgSerProGlyGlyHisSerAspGlyAlaSerProArgSerProValProThrThrLeu 124
 QY 1017 -----TGAGCATCCCCCGCCCAAGTCTCTGTCAGACCTAAGC 979
 DB 125 ProGlyLeuArgHisAlaProTrpGlnGlyProArgGly-----ProProAspSer 141
 QY 978 TGAGAGAACACTCCAGATGATGATTGCTCCGACGAG-----CCC 940
 DB 142 ProAspGlySerProLeuThrProValProSerGlnMetProTrpLeuValAlaSerPro 161
 QY 939 CCAACCGCTGCTGTGATGTGACAGCAGCAGCAGGAGGCTTATGATTCATCTCACAC 880
 DB 162 GluProPro-----GlnSerSerProThrProAlaPheProLeuAlaAlaSer 177
 QY 879 TATTGAACACAGTGAATGGGCTGCCA-----CGAACCTCCGAGGAG 835
 DB 178 TyrAspThrAsnGlyLeuSerGlnProProLeuProGluLysArgHisLeuProGlyPro 197
 QY 834 GGGCAGACGCCCC---CAGTGGGAGAGTACCTGAGCTTTCAGCCAGCCAGGAGTGG 778
 DB 198 GlyGlnGlnProGlyProTrpGly-----ProGlnGlnAlaSerSerProAla--- 213
 QY 777 AGAGGCGATCATACCCCATGATTCCTGACTAGAGCAGCAGAGTCCCATAGCTAGAGCTG 718
 DB 214 ArgGly----- 215
 QY 717 GTAACACTCATATTTAACATTGGCATTATGACAGTTTGCTCTGATCATCTCTGT 658
 DB 216 -----IleSerHisIleValThrPheAlaProLeuLeuSer----- 227
 QY 657 CTGGGATGTTAGCTTTGTCTGACGGGCTCCAGGTACATGACAGTCTGCCAGTGCAG 598
 DB 228 -----Asp 228
 QY 597 TCATCCCTCCACAGCAGCAGCCAGCTGTGAGATAGTCACTGTCGACCGACCTC 538
 DB 229 AsnValProGlnThrProGluProProThrGlnGlnSerGlnSerAsnValLysPheVal 248
 QY 537 AGATAGGTCAACGACGAGCTTCAATATGTGTCTTTGGTATCAGGACCACTTGGC 478
 DB 249 GlnAspThrSerLys-----PheTrp-TyrlsPheHisLeu 261
 QY 477 TCCATCC-----AGTCTTGAAGAAAGACAGCTGCCACCTGGGGGGCAGAGGTGAG 427
 DB 261 TyrAspGlnAlaIleAlaLeuLeuLysAspLysAspPro----- 274
 QY 426 AGGTAAACAGGCTGATGTTCTCTTCTGAGGACAGACCATCTCCCTAATTGACTG 367
 DB 275 -----GlyAlaPheLeuIle---ArgAspSerHisSer----- 284
 QY 366 CTTGAAGACAGCATTAACCTCTCTCATCATCAGACACCCCTCTGACTTCGAGTGG 307
 DB ----- 307

DB 285 -PheGlnGlyAlaTyr-GlyLeuAlaLeuLysValAlaThrPro----- 298
 QY 306 GGATCACTATCTATCAACCTCCAGAGAGGTCTGCATCTCCCTGGAAGGC----- 254
 DB 299 -----ProProSerAlaGln-----ProTrpLysGlyAspPro 310
 QY 253 -----TTCCTCCACTGGCTC-----ACCTGTCAGGTGT-----CTGATG 217
 DB 310 AlGlnGlnLeuValArgHisPheLeuIleGluThrGlyProLysGlyValLysIleLys 330
 QY 216 GTGACCAACAAAGCCCGACACTGGC-----ATGCTCTACACAGCATCCACAGCTCA 163
 DB 330 LysProSerGluProTyrPheGlySerLeuSerAlaLeuValSerGlnHisSerLys 350
 QY 162 GCCAGCT----- 145
 DB 350 erProIleSerLeuProCysCysLeuArgIleProSerLysAspProLeuGlnGluThrP 370
 QY 144 AGAGGACACTCCCTGCTGTGATGAGTGCACAGCAGGATGCTTCTGGGGAGTGTG 85
 DB 370 roGlu---AlaProValProThrAsnMetSerThrAlaAlaPheLeuArgGlnGlyA 389
 QY 84 CTCTCTGGATACACTGTGACTC 62
 DB 389 laAlaCysSerValLeuTyrLeu 396
 RESULT 13
 US-10-176-847-68
 ; Sequence 68, Application US/10176847
 ; Publication No. US20030068636A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Velby, Pectet Ole
 ; TITLE OF INVENTION: COMPOSITIONS, KITS, AND METHODS FOR
 ; TITLE OF INVENTION: IDENTIFICATION, ASSESSMENT, PREVENTION, AND THERAPY OF BREAST
 ; FILE REFERENCE: MRI-039
 ; CURRENT APPLICATION NUMBER: US/10/176,847
 ; NUMBER OF SEQ ID NOS: 112
 ; SOFTWARE: FastSeq for Windows Version 4.0
 ; SEQ ID NO 68
 ; LENGTH: 1236
 ; TYPE: PRT
 ; ORGANISM: Homo sapiens
 US-10-176-847-68
 Alignment Scores:
 Pred. No.: 0.186 Length: 1236
 Score: 120.50 Matches: 88
 Percent Similarity: 32.37% Conservative: 35
 Best Local Similarity: 23.16% Mismatches: 135
 Query Match: 4.71% Indels: 122
 DB: 9 Gaps: 19
 US-09-989-919-15 (1-1397) x US-10-176-847-68 (1-1236)
 QY 1199 GGCAGAAAGCCCGGCGCCACCCCG-----GCCCGAGTCTG 1161
 DB 926 GlyArgAlaAlaProProProSerProCysProThrProArgSerGlyProSerProCys 945
 QY 1160 GCTCAACAC-----ACACTGCTCCCTGACCCAGCTTTGG----- 1125
 DB 946 LeuProThrProAspProProProGluProSerProThrGlyTyrGlyProProAspGly 965
 QY 1124 -----CTATGCTGAGGTGTGCACT-----CTGCCCCAGACCCCTGG 1086
 DB 966 GlyArgAlaAlaLeuValArgAlaProGlnProGlyArgProProThrProGly 985
 QY 1085 CTGGCTGGAGTGTCCCTGCTGGAGGTACAGATACCTCCAGGTACAGATCAC 1026
 DB 986 -----ProProLeuSerAspValSerArg-----ValSerArgArg 997
 QY 1025 CCACATCTGAGCATCCCCCGCCCAAGTCTCTGTCAGACAC---TAACCTGAGAGACA 969

OY	1141	CCTGACCCCAAGCTCTTGCTCAATGCTAGAGTGTGCAACCTCTGGCCCT	1098
Db	1805	ProtaGProGln-ProProProProGlnSerCysAlaLeuProProtaGSerLeuPr	1824
OY	1095	----GACCCCTGGGGGTGGCTGGAGATGTCCCTCTCGGAGAGGTACAGATAGCCTCCCC	1040
Db	1824	oSerAspProPheSerArgValProValSerProGlnSerGlnSerSerGlnSerPr	1844
OY	1039	AGGTACAGATATCACCACATCTCTGAGCATCCCCGCCAAGTCTCTGTCCAGACTTAAG	980
Db	1844	oleuThrProArgProLeuSer-----AlaGluAlaPheCysProSer-----	1858
OY	979	CTGAGAGAAACCTCCACAGATGATGTCTCCACAGGAGGCCACCGCCCTGTGTGGCATG	920
Db	1859	----ProValThrProArgPhe-----GlnSerProAspPro-----	1865
OY	919	TGCAGACAGCAGCAGAGGGGCTTTAGCTATTCACACTTAATGAACACAGTGAATGGG	860
Db	1870	-----TyrSe	1871
OY	859	CTGCCACGGAACCGTCCAGGAGGGG-----CACAGCCCCCAGT	818
Db	1871	rArgProProSerArgProGlnSerArgAspProPheAlaProGlnHisLysProProAr	1891
OY	817	GGGAGACTAGCTGAGCTTGTCTTACAGCCACAGGGCTTGGAGAGGCACT-----CATAC	764
Db	1891	gProGlnProProGluValAlaPheLys-----AlaGlySerLeuAlaHisThr	1907
OY	763	CCCCATG---ATTCCGTGACTAGAGGACGACAGTCCCATAGCTAG--GCTGTGAATAC	708
Db	1907	rSerLeuAlaGlyGlyPheProAlaAlaLeuProAlaGlyProAlaGlyGluLeu-	1926
OY	707	ATTTAACTTTGGATATTGACAGTTTGTCTGTGATCATCTGTCTGTGGAGATGT	648
Db	1927	-----HisAlaLysValProSerGlyGlnProProAsnPheValaLysSerProGlyTh	1944
OY	647	AGCTTTGCTCTGAGGGGTTCCCAAGTGACATGCAGTGTCCCAAGTGCAAGTCACT---	593
Db	1944	rGlyAlaPheValGlyThrProSer-----PrometaArgPheThrPhePr	1959
OY	592	-----CCCTCACAGGACGAGCCAGCTG	570
Db	1959	oGlnAlaValGlyLupProSerLeuLysProProValProGlnProGlyLeuProProPr	1979
OY	569	CTGAGATAGTCAAGTGTGT---CAAGCCAGCCCTCAGATAGTACAGCCAGGCTTCAAC	513
Db	1979	oHisGlyLysLeuSerHisPheGlyProGlyProThrLeuGlyLysProGlnSerThrAs	1999
OY	512	ATATGTGATCTTTGTTGGATCAGGGACCACTTGGCTCATCACTGTTGAAAAAGACAG	453
Db	1999	nTyThrValAlaThrGlyAsn-----PheHisPro-----Se	2010
OY	452	CTGCCGACCTGGGGGCGAGAGTGAAGSTAAACAGGCTGATGTTTCTTTCTGGAGG	393
Db	2010	rGlySerProLeuGlyProSerSerGlySerThrGlyGlnSerTyrglyLeuSer-ProL	2030
OY	392	CAAAACATTTCTCCCTATTTGACTGCTTGAGAACAGATTAACCTCTCTCACTCACT	333
Db	2030	euaArgProProSerValLeu-----	2036
OY	332	CAGCACCCCTCTACTCTCCAGTGGGGATCACTATCTATCACTCCAGAGAGTCTCTG	273
Db	2037	----ProProProAlaProAsp---GlySerLeuProTyL---LeuSerHisGlyLys	2053
OY	272	CCATCTCCCTTGAAGGGCTTCTCCCACT-----G	243
Db	2053	erGlnArgSerGlyLysThrSerProValGlnLysArgGluAspProGlyLThrGlyMetG	2073
OY	242	GCTCAACCTGTCAGAGTGTGTGTTGGTGAACCAAAAGGCCGACACTGGGCAAGGGCTCT	183
Db	2073	lySerSerLeuAlaThrValaGluLeuProGlyLThrGlnAspProGlyLysSerGlyLeuS	2093

QY 182 CACAGACATCCACAGCTCAGCCAGGTGAACACTGAAGAGGACGACCTC 133
Db 2093 egiInThrGluLeuGlnIyGlnArgGlnArgLeuArgGluLeu 2109	

RESULT 2

ALR protein - human
 C:Species: Homo sapiens (man)
 C:Date: 24-Mar-1999 #sequence_revision 24-Mar-1999 #text_change 21-Jul-2000
 C:Accession: T03454
 R:Prasad, R.; Zhadanov, A.B.; Sedkov, Y.; Bullrich, F.; Druck, T.; Rallapalli, R.; Yano
 Oncogene, 15, 549-560, 1997
 A:Title: Structure and expression pattern of human ALR, a novel gene with strong homolo-
 A:Reference number: Z14954; MUID:97388474; PMID:9247308
 A:Accession: T03454
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: mRNA
 A:Residues: 1-5262 <PRA>
 A:Cross-references: EMBL:AF010403; NID:G2358284; PIDN:AAC51734.1; PID:G2358285
 C:Genetics:
 A:Gene: ALR
 A:Map position: 12
 C:Superfamily: human ALR protein
 C:Keywords: alternative splicing

Alignment Scores:

Pred. No.:	0.0291	5628
Score:	127.50	Length:
Percent Similarity:	33.97%	Matches:
Best Local Similarity:	25.33%	Conservative:
Query Match:	4.98%	Mismatches:
DB:	2	Indels:
		Gaps:

US-09-989-919-15 (1-1397) X T03454 (1-5262)

Qy	1219	CTTCACTGAAGCAAAATGAG-----	-GCAGAAAGGCCCGGCC	1181
Db	2070	LeuGlyLeuAarProGInGluPrProProAlaGlnAlaLeuAlaProSerProProSer	2089	
Qy	1180	CACCCC-----CGCCCCAGGCTCTGGCTCAAAACACAC-----	CTGCTC	1142
Db	2090	HisProAspIlePheArgProGlySerTyrThrAspProTyrIleGlnProProLeuThr	2109	
Qy	1141	CGTAGCCCCAGCTGTGGCTCATGCTAGAGGTGGCCACTCTGCCCT-----	1096	
Db	2110	ProArgProGln--ProProProProGInSerCysGlyAlaLeuProProArgSerLeuPr	2129	
Qy	1095	-----GACCCCTGGAGCTGGCTGGAGAGTGTCCCTGTGGAGGGTCAGAGATAGCTTCCC	1040	
Db	2129	oSerAspProPheSerArgValProValSerProGInSerGlnSerSerGlnSerPr	2149	
Qy	1039	AGGTACAGAAATCACCCACATCTCTGAGACATCCCCCGCAAGTCTCTGTCCAGACTAAG	980	
Db	2149	oleuThrProAixProLeuSer-----AlaGlnAlaPheCysProSer-----	2163	
Qy	979	CTGAGAGAACACTCCACAGATGATTTGCCCCAGCAGGCCCCACCGCTGCTGGCATG	920	
Db	2164	-----ProValThrProArgPhe-----GlnSerProAspPro-----	2174	
Qy	919	TGCAGCAGCAGCAGCAGAGGGGCTTAGCTACATTCTCACACTATTGAAACACAGATGATGGG	860	
Db	2175	-----TyIse	2176	
Qy	859	CTGCCACCGAACCCTGCCAGGGAGGG-----CACAGCCCCCAGTG	818	
Db	2176	IaArgProPheSerArgProGInSerIaArgAspProPheAlaProLeuHisLysProProAr	2196	
Qy	817	GGGAGACTAGCCTGAGCTTGCTTACAGCCACACAGGGGTGGAGAGGAGT-----CATAC	764	
Db	2196	gProGInProProGluValAlaPheLys-----AlaGlySerLeuAlaHisThr	2212	
Qy	763	CCCCATG--ATTCTTGAGTAGAGAGCAGACAGTCCCATAGCTAG--GCTGGTAAACTCAC	708	

Db 2212 rSerLeuGlyAlaGlyPheProAlaAlaLeuProAlaGlyProAlaGlyLeu-- 2231
 QY 707 ATTTAAACATTGGCATTTATGACAGCTTTGCTGGTGCATCTGTCTGGAGATGT 648
 Db 2232 -----HisAlaValSerGlyGlnProProAsnPheValArgSerProGlyTh 2249
 QY 647 AGCTTTGTCTGACAGGGGTTCCAGTGACATGACAGTGTCTGCGAGTCACT-- 593
 Db 2249 rGlyAlaPheValGlyThrProSer-----ProMetArgPheThrPheP 2264
 QY 592 -----CCCTCCACAGGCAAGCCAGCTG 570
 Db 2264 oGlnAlaValAlaGlyLeuProSerLeuysProProValProGlnProGlyLeuProProP 2284
 QY 569 CTGAGGATAGTACAGCTGCT--CAGCCGAGCCCTCAGATAGTACAGGAGGCTTAC 513
 Db 2284 oHisGlyIleAsnSerHisPheGlyProGlyProThrLeuGlyLeuProGlnSerThrAs 2304
 QY 512 ATATGTGCTTTGTGGATCAGGAGACCACTGCTTCATCCAGTTTGAAAAAGACAG 453
 Db 2304 nTyThrValAlaThrGlyAsn-----PheHisPro-----Se 2315
 QY 452 CTGCCACCTGGGGGGGAGAGGTGAGAGTAAACAGGCTGATGTTCTTTCTGGAG 393
 Db 2315 rGlySerProLeuGlyProSerSerGlySerThrGlyLeuSerTyGlyLeuSer--ProL 2335
 QY 392 CAAGACCATCTCCCTAATTGACTGCTTGAAGACAGTATACCTCTCTCATCACT 333
 Db 2335 euArgProProSerValLeu----- 2341
 QY 332 CAGACACCCCTCTGACTTCCGATGGGAGTCACTATCTATCACTCCAGAGGCTCTG 273
 Db 2342 -----ProProProAlaProAsp--GlySerLeuProTy--LeuSerHisGlyAlaAs 2358
 QY 272 CCAATCCCTTGAAGGGCTTCTCCACT-----G 243
 Db 2358 eArgInArgSerGlyIleThrSerProValGlyLysArgLysAspProGlyThrGlyMetG 2378
 QY 242 GCTCACCTGTCAGGTGCTGCTGGTGTGACACAAAGGCCGACACTGGGAGATGGCTCT 183
 Db 2378 lYerSerLeuAlaThrAlaPheGlyLeuProGlyThrGlnAspProGlyMetSerGlyLeuS 2398
 QY 182 CACAGACATCCACAGCCTCAGCAGCTTGAACACTGAAGAGAGCAGCTC 133
 Db 2398 eArgInThrGlyLeuGlyLysGlnArgLysInArgLysArgLeuArgGlyLeu 2414

RESULT 3

149705 glutamate receptor channel subunit epsilon 3 - mouse
 C:Species: Mus musculus (house mouse)
 C:Date: 02-Aug-1996 #sequence_revision 02-Aug-1996 #text_change 21-Jan-2000
 C:Accession: I49705
 R:Kutsuwada, T.; Kaahiwabuchi, N.; Mori, H.; Sakimura, K.; Kushiya, E.; Araki, K.; Meguri
 Nature 359, 36-41, 1992
 A:Title: Molecular diversity of the NMDA receptor channel.
 A:Reference number: 149704; PMID:92310564; PMID:1377365
 A:Accession: I49705
 A:Status: preliminary; translated from GB/EMBL/DBD
 A:Molecule type: mRNA
 A:Residues: 1-1239 <RES>
 A:Cross-References: GB:D10694; NID:G538239; PIDN:BA01536.1; PID:G538240
 C:Superfamily: N-methyl-D-aspartate receptor 2C; glutamate receptor homology
 C:Keywords: neurotransmitter receptor
 F:425-852/Domain: glutamate receptor homology <GRH>

Alignment Scores:

Pred. No.: 0.0388 Length: 1239
 Score: 126.00 Matches: 94
 Percent Similarity: 30.61% Conservative: 23
 Best Local Similarity: 24.61% Mismatches: 112
 Query Match: 4.92% Indels: 153
 DB: 2 Gaps: 24

US-09-989-919-15 (1-1397) x I49705 (1-1239)
 QY 1196 AGAAGAGCCCGGCGCCGACCCGCGCCAGGCTGCTGAACACACCTGCTCCCTTA 1137
 Db 927 ArgAlaPheProAlaPro-----ThrThrSerGlyPro-- 937
 QY 1136 CCCCAGCTTGGGCTCACTGCTGAGCTGACCTTCCCTGACCCCTGAGC----- 1086
 Db 938 -----ArgSerCysThrProGlyPro-----ProGlyGlnProSer 949
 QY 1085 ---CTGGCTGGAGAGTCCCTGCTGGAGGTCAGAGATAGCTTCCCAAGTACAGAT 1029
 Db 950 ProSerGlyTTArgProPro-----GlyGlyGlyArg--ThrProLeuAlaArg 966
 QY 1028 CAGCCACATCTGGAGACATCCCGGCGAAG-----TCTCTCTCCAGACTTAAGT 978
 Db 967 AlaProGlnProProAlaArgProAlaThrCysValGlySerProGlnProAlaValSer 986
 QY 977 GAGAGAACACTCCACAGATGATGTTCCSAGCAGCCCGCCGCTGTGTGSCATGTG 918
 Db 987 ArgAlaSerCysArgHisAlaThrAspAlaArg-----ThrProVal 1000
 QY 917 CAGCAGACAGCAGAGGGCTTTAGTACATTTCCACATTTGAACACACTAGTGGCCT 858
 Db 1001 ArgValGlyHisGlnGly-----SerHisLeuSerAlaSer 1012
 QY 857 GCCCAGAGAACCGTCCAGAGAGGGGAGCAGAGCCCGCAGTGGGAGACTAGCTGAGCTTG 798
 Db 1013 GlyArgArgAlaLeuProGlyArg-----SerLeu 1022
 QY 797 CTTCAGCCCGAC-----AGGGTTGGAGAGGAGCAGTC 768
 Db 1023 LeuHisAlaHisLeuSerHisGlySerSerPheProAlaGlyLysArgGlyArgProPhe 1042
 QY 767 ATACCCCATGATTTCTGACTAGAGCCAGCA-----GTCCCATAGCTAGGC 720
 Db 1043 LeuPro-----LeuPheProGlnProProGlnProAlaAspArgLeuLeuGly 1059
 QY 719 TGGTAACATCAATTTAATTTGACATTTATGACAGTTTGCTGGTGCACATCTGTGT 660
 Db 1060 ---ProGlyGlnLeuAlaArgArgGlyAlaLeuLeuAlaAlaThrAla----- 1075
 QY 659 GTCTGGAGATTTAGCTTTGTCTGTCAGAGGGTCCAGG---TGACATGACAGTGTGACAT 603
 Db 1076 -----ArgGly--ProArgProArgHisAlaSerLeuProSe 1087
 QY 602 GCAAGTC-----ACTCCCTCCACAGGCAAGCCAGC---TGTGAGGATAGTC 558
 Db 1087 rSerValAlaGlyAlaPheThrArgSerAsnProLeuProAlaArgCysThrGlyHisAl 1107
 QY 557 AGCTGGTACAGCCGAGCCCTCAGATAGTGTAGCCAGAGGCCCTTACATGTGTCTTGT 498
 Db 1107 AcCysAlaCysProCysProGln----- 1114
 QY 497 GGGATCAGGAGACACTTGGCTCCATCAGCTTTGAAAAAGACAGCTGGCGA----- 446
 Db 1115 -----SerArgProSerCysArgHisValAla 1123
 QY 445 -----CTGGGGGGCAGAGGTGAGAGTAAACAGGCTGATGT 408
 Db 1123 acGlnThrGlnSerLeuArgLeuProSerGlyArgGlyAlaCysValGlyGlyAlaProAl 1143
 QY 407 TTCTCTTTGTGAGGAGGAACCATTTCTCCCTAATTGACTGCTTGAAGACAGTATAC 348
 Db 1143 acGlyValAlaAla-----ThrTrpGlnPhe 1151
 QY 347 TCTCTCTCAT-----CAGCTCAGACCCCTGACTTCCGATGGGAGTACTAT 297
 Db 1151 oArgGlnHisValCysLeuHisHisThrHisLeuProPheCysTrpGlyThrValCys 1171
 QY 296 CTATCAACCTCCAGAGAGGCTCTGCATCTCCCTTGGAAAGGCTTCCCATGAGCTCAC 237
 Db 1171 sarGHisProProProCysSerSerHisSerProTrp----- 1183

QY 236 CTGTCCAGTGTGCTGTGTGTGT-----GACCACAAAGCCCGCACTGGG 192
 Db 1184 -----LeuilegilyThrtprgluProProSerHisArgIlyrThrglu 1199
 QY 191 CATG 188
 Db 1199 yLeu 1200

RESULT 4
GNMVR4

structural polyprotein - rubella virus (strain RA27/3 vaccine)
 N:Contains: capsid protein; membrane glycoprotein E1; membrane glycoprotein E2
 C:Species: rubella virus
 C>Date: 30-Sep-1990 #sequence_revision 30-Sep-1990 #text_change 16-Jul-1999
 C/Accession: S04800
 R:Nakhasi, H.L.; Thomas, D.; Zheng, D.; Liu, T.Y.
 Nucleic Acids Res. 17, 4393-4394, 1989
 A:Title: Nucleotide sequence of capsid, E2 and E1 protein genes of rubella virus vaccine
 A:Reference number: S04800; PMID:89296505; PMID:2740235
 A:Accession: S04800
 A:Molecule type: mRNA
 A:Residues: 1-1063 <MAX>
 A:Cross-references: GB:X14871; NID:961915; PID:CAA33016.1; PID:961916
 A>Note: the authors translated the codon CGC for residue 207 as Val and AGC for residue
 C:Superfamily: rubella virus structural polyprotein
 C:KeyWords: capsid protein; glycoprotein; polyprotein; transmembrane protein
 F:1-300/Product: capsid protein C #status predicted <NPC>
 F:278-298/Domains: transmembrane #status predicted <TM1>
 F:301-582/Product: transmembrane glycoprotein E2 #status predicted <E2>
 F:559-555/Domains: transmembrane #status predicted <TM2>
 F:583-1063/Product: membrane glycoprotein E1 #status predicted <E1>
 F:1032-1050/Domains: transmembrane #status predicted <TM3>
 F:353,371,410,429,658,759,791/Binding site: carbohydrate (Asn) (covalent) #status predicted

Alignment Scores:

Pred. No.: 0.131 Length: 1063
 Score: 119.50 Matches: 106
 Percent Similarity: 31.33% Conservative: 35
 Best Local Similarity: 23.56% Mismatches: 143
 Query Match: 4.67% Indels: 166
 Db: 1 Gaps: 24

US-09-989-919-15 (1-1397) x:GNMVR4 (1-1063)

QY 1184 GCGCCACCGCCCGCCAGCTCTGCTCAAC-----ACACCTGCTCCCTGACC 1134
 Db 77 AlarProProProgluIngluArgIngluSerArgSerInthrProAlaProLysPro 96
 QY 1133 CAGCTTGAGCTAGCTGAGGTGACCTTGGCCCTGACCCCTGGGGCTGGGGAG 1074
 Db 97 SerArg-----AlaPro-----ProgluIngluProglu----- 105
 QY 1073 TGTCCCTGTGCGAGAGTACAGATAGCTCCCGAGTACAG-----AAT 1029
 Db 106 -----ProProArgMetGlnThrArgIlyrArgIlySer 116
 QY 1028 CACCCACATCTCTGAGAGATCCCGCCAGTCTCTCTGTCAGACCTAGAGAGA--- 972
 Db 117 AlarProArgProgluLeuGlyProProThrAsnProPheGlnAlaAlaValArgIly 136
 QY 971 -----ACACTCCACAGATGATGTGCCCAAGCAGCCCGCCAGCTCTCT 926
 Db 137 LeuArgProProLeuHisAspPro-----AspThrGlu-AlaProThrGluAlaCysValTh 155
 QY 925 -----GGCATGTGCGAGAGCAGACAGAGGGCTTAGCTACACTTCACACTATTGAAC 870
 Db 155 rSerTtrpLeuTrpSerGluGlyGlyAlaValAlaPheTyrArgValAspLeuHisPheTh 175
 QY 869 AGTATGAGGGGTGCGCCACGGAACCGTCCAGAGA----- 836
 Db 175 rAsnLeuGlyThrProProLeuAspGluAspGlyArgTrpAspProAlaLeuMetTyrAs 195

QY 835 -----GGGGACAGCCCGCCAGTGGGAGACTAGACCTGAGCTTGACAGCCACAGG 783
 Db 195 nProCysGlyProgluProPro-----AlaHisValAlaArgAlaTyrAsnGlnProAlaG 214
 QY 782 GTTGGAGAGCAGATACATACCCCCAGATGATTCCTGATAGAGCCAGCAGACTCCCATAGCTA 723
 Db 214 yAspValAlaArgIlyValTtrpGlyLysGlyGlyGlnArgThrTyrAlaGluIngluAspPheArg 234
 QY 722 GCGGTGAACATCACTTTAACTTTGAGCATTTATGAGCATTTATGAGCATTTATGAGCATTT 663
 Db 234 lGlyGly-----ThrArgTtrpHisArgPheLeuArgMetProValArgGly----- 249
 QY 662 TCTGTCTGGAGATGATAGCT-----TTCCTCAGAGGGTTCCAGATGATGATGATGATGCC 606
 Db 250 -LeuArgGlyAspThrAlaProLeuSerPro-----HisThrThrGluArg 264
 QY 605 AGTGCAGATGATCTCTCCACAGAGAGCCAGCTGCTGAGATGATGATGATGATGATGATGATG 546
 Db 264 glileGlu-----ThrArgSerAlaArgHisProSer----- 274
 QY 545 GAGCCCTCAGATAGTACAGCCAGAGCCCTTTCACATATGCTGTGTTGGATCAGAGGAC 486
 Db 275 -SerIleArgPheGlyAlaProGlnAlaPheLeuAlaGlyLeuLeuAlaValAla 294
 QY 485 CACTTGGCTTCATCCAGCTTGTGAAAAGACAGCTGCCAGCTGGGGGAGAGGTGAGA 426
 Db 294 a-ValIlyThrAlaArgAlaGlyLeuGlnProArgValAspMet----- 308
 QY 425 GGTAAAGAGGCTGATGATTTCTTTCTTGGAGGACAGACATTTCCCTAATTGATGCTG 366
 Db 308 ----- 308
 QY 365 CTGGAAGACAGATATCTCTCTCTCATACCTCAGACACCCCTGATCTCCGATGGG 306
 Db 309 -----AlaAlaProProMet----- 313
 QY 305 GATCACTATCTATCAACCTCAGAGAGCTGCTGCCCTGCTGGAAGAGGCTTCCCA 246
 Db 314 -----Pro-ProGlnPro-----Pro 318
 QY 245 CTGGCTCAGCTGCTCAGATGCTGTGTTGATGATGATGATGATGATGATGATGATGATGATG 186
 Db 319 ArgAlaHisGlyGlnHisTyrGlyGlnHisTyrGlnHisTyrGlnHisTyrGlnHisTyrGln 338
 QY 185 TCTCAGACATGTCACAGCTCAGCCAGCTGTAACACTGAAGAGAGGAGCTCCCTGCT 126
 Db 339 gLynHis-----GlyGlyThrLeuArgVal----- 347
 QY 125 GGTAGAGTGGCCAGCAGCGGTAGT-----CTTCTGGGGGATG 87
 Db 348 GlyGlnHisHisArgAsnAlaSerAspValLeuProGlyHisTyrLeuGlnGlyTtrp 367
 QY 86 TGCTGTCTGGA-----TACACTGATCTGCTACTGC----- 54
 Db 368 GlyCysTyrAsnLeuSerAspTtrpHisGlnGlyThrHisValCysHisThrIlyHisMet 387
 QY 53 -----TGCTTG-----CCGTGAGCTTCTGACAGAT 30
 Db 388 AspPheTtrpCysValGlnHisAspArgProProProAlaThrProThrSerLeuThrThr 407
 QY 29 ACTGCCCGCTCCGGTACAGGTGCA 6
 Db 408 AlaAlaHisSerThrThrAlaAla 415

RESULT 5

GNMVR77

structural polyprotein - rubella virus (strain HPV77)
 N:Contains: capsid protein; membrane glycoprotein E1; membrane glycoprotein E2
 C:Species: rubella virus
 C>Date: 30-Jun-1991 #sequence_revision 30-Jun-1991 #text_change 16-Jul-1999
 C/Accession: J00087
 R:Zheng, D.; Dickens, L.; Liu, T.Y.; Nakhasi, H.L.
 Gene 82, 343-349, 1989

US-09-989-919-15 (1-1397) x A39255 (1-897)

QY 1221 GGCTTCATTGAAAGCAATAGAGCAGAAAGCCCGCCAGCCCGCCCGAGGTCTCT 1162
 DB 464 GYIATYGLYrtrgleuAarglystrpGLuLylleProanProserLys 483
 QY 1161 GGC-----TCMAACCACTGCTCCCTGACCCCACTTGGCTCATGCTAGAG 1111
 DB 484 SerHisleuPheGlnsnGlySerAlaGlnLeuTrpProGlnLysSerMetSerAlaPhe 503
 QY 1110 TGACAGCTGCGCCCT-----GACCCCTGGGCTGCGCCTG-----GGA 1075
 DB 504 ThrSerGlySerProHisGlnGlyProTrpGlySerArgPheProGlnLeuGlnGly 523
 QY 1074 GTGTCCCTGTC-----GGAGAGTCAGAGATAGCCTCCCGAGTACAGAAATCAGCCACA 1021
 DB 524 ValPheProValGlyPheGlyAspSerGluVal-----SerProLeu 537
 QY 1020 TCCTGGAGCATCCCGCCCAAGTCTCTGTCCAGACTTAAGCTGAGAGAACTCCACGA 961
 DB 538 ThrIleGlnAspProLysHisValCysAspProProSerGlyProAspThrThrProAla 557
 QY 960 TGAGATT-----GTCCCGCAGC-----CAGCCCGCCAGCCGCTGCTGAGCA 922
 DB 558 AlaSerAspLeuProThrGlnGlnProProSerProGlnProGlyProProAlaAlaSer 577
 QY 921 TGTGACAGCAGCAGCAGCAGCGGCTTAGCTA----- 891
 DB 578 HisThr-ProGlnLysGlnAlaSerSerPheAspPheAsnGlyProTyrLeuGlyProPr 597
 QY 890 -CATTCACACATTTGAACACACTGATGGGGCTGCCAGGAACTCCCGAGGAGGG 832
 DB 597 OHAsSerArgSer-LeuProAspIleLeuGlnGlnProGlnProGlnGlnGlyLys 617
 QY 831 CACAGC-----CCCCAGTGGAGACTAGCTGAGCTTACAGCCCAAGAGGTGG 778
 DB 617 ertGlnLysSerProProGlnGlySerLeuGlnLysLeuGlySerLeuProAlaGlyLys 636
 QY 777 AGAGGCGATCATACCCCATGATTCCTGAGCTAGAGCCAGCAGTCCCATAGTAGAGCTG 718
 DB 636 InValGlnLeuValPro----- 641
 QY 717 GTAACCTCACATTTTAACATTGGCATTTATGACGTTTGTCTGTCATCTGTCTGT 658
 DB 641 ----- 641
 QY 657 CTGGATGATGATCTTGTCTGACAGGGGTCCAGAGTGACATGACAGTCCAGTGCAG 598
 DB 642 -----LeuAlaGlnAlaMetGlyProGlnGln-----AlaValGlnValGln 656
 QY 597 TCACCTCCCTCCACAGGAGCCAGCTGCTGAGAGATGACAGTGTGTCACCCAGCCCTC 538
 DB 656 rgaArgProSerGlnGly-----AlaAlaGlySerProSerLeuG 669
 QY 537 AGATAGCTCAGCAGAGAGCCTTACATATGTGTCTTGTGGATCAGGAGACACTTGGC 478
 DB 669 IuSerGlyGly-----GlyProAlaProProAla 679
 QY 477 TCATCAGCTTGAAGAAAGACAGCTGCGAAGCTGGGGGAGAGAGTGAGTAACAG 418
 DB 679 euGlyPro-----ArgValGlyGlyGlnAspGlnLysAspSerP 692
 QY 417 GGTGATGATGTTCTTCTTGGAGGCAAGACCATTTCTCCATATTGACGCTTGAGA 358
 DB 692 roValAlaIlePrometSerSerGlyAspThrGlnAspProGlyValAlaSerGlyTyr 712
 QY 357 CAGGTATACCTCTCTTCATCAGCTCAGACACCCCTGTGACTTCCAGTGGGGATCACTA 298
 DB 712 alSer-SerAlaAspLeuVal-----PheThrProAsnSerGlyAlaSer 726
 QY 297 TCTATCAAGCTC-----CAGAGAGGTCTGCAGATCTCCCTT 262

DB 727 SerValSerLeuValProSerLeuGlyLeuProSerAspGlnThrProSerLeuGlyPro 746
 QY 261 GGAAGGCTTCT-----CCCACTGGCTCAGCTGTCCAGTGC----- 226
 DB 747 GlyLeuAlaSerGlyProProGlnLysAlaProGlyProValLysSerGlyPheGlnGlyTyr 766
 QY 225 GTCTGTTGGTGTACCAACAAAGGCGCAGACCTGGGATGGCTCTCAGACATCCACAGCC 166
 DB 767 ValGlnLeuProProIleGlnGlyTyr-SerProArgSerProAlaGlnAsnProValPr 786
 QY 165 T-----CAGCAGGTTGAACACTGAAAGAGAGCAGCTCCG 130
 DB 786 ProGlnAlaLysSerProValLeuAsnProGly-----GluArg 799
 QY 129 TGGTGTAGATAGCCAGCAGCGGTAAGTCTTCTGGGGATGATGCTGTCTGGATAC 70
 DB 799 ProAlaAspValSerProThrSerProGlnProGlnGlyLeuValLeuGlnGln 819
 QY 69 TGTACTCGTACTGCTGCTGCTCCG 45
 DB 819 IGLYAsp---TyrCysPheLeuPro 826

RESULT 7

T45025
 Mucin MUC5B, tracheobronchial [Imported] - human (fragment)
 C/Species: Homo sapiens (man)
 C/Date: 21-Jan-2000 #sequence_revision 21-Jan-2000 #ext_change 21-Jul-2000
 C/Accession: T45025
 R/Designer: J.L.; Guyonnet-Duperrat, V.; Porchet, N.; Aubert, J.P.; Laine, A.
 J. Biol. Chem. 272, 3168-3178, 1997
 A>Title: Human mucin gene MUC5B, the 10.7 kb large central exon encodes various alternat
 A/Reference number: Z22899; MUID:97166151; PMID:9013550
 A/Accession: T45025
 A/Status: preliminary; translated from GB/EMBL/DBJ
 A/Molecule type: DNA
 A/Residues: 1-3570 <DBS>
 A/Cross-references: EMBL:Z72496; NID:G1834502; PIDN:CAA6577.1; PID:G1834503
 A/Experimental source: placenta
 C/genetics:
 A/gene: MUC5B

Alignment Scores:

Pred. No.: 0.172 Length: 3570
 Score: 118.00 Matches: 100
 Percent Similarity: 31.768 Conservative: 55
 Best Local Similarity: 20.498 Mismatches: 159
 Query Match: 4.61% Indels: 174
 DB: 2 Gaps: 19

US-09-989-919-15 (1-1397) x T45025 (1-3570)

QY 1294 CCAACAAGATGACAAAGATTCTGATATTCAAACTCAGAGAGAGGAGAAAGCTGTGTT 1235
 DB 795 ProSerSerThrGlnThrSerGlyThrProProSerLeuThrThrThrAlaThrThr 814
 QY 1234 TTGGCTGCTTTAGAGCTTTCATTGAAGCAATAGAGCAGAAAGCGCGGCCCAAGCC 1175
 DB 815 ILeThrAla-----ThrLysThrThrAsnProSerSerThr-ProGlyThrThrPr 832
 QY 1174 CGCCCGAGCTCTGAGCTCAAAACACACTGTCTCCCTGACCCAGTGGCTCAGTCAAGCTGA 1115
 DB 832 oIlePro-----ProValLeuThr----- 838
 QY 1114 GGTGTACACTCTGCGCCCTGACCCCTGGGCTGAGCTGGAGAGTGTCCCTGTGGAGAGTC 1055
 DB 839 -----ThrThrAla-----ThrThrProAlaAlaThrSe 848
 QY 1054 AGAGATAGCTCCCGCAGTAACAGATCACACATCTGGAGATCCCGCCAGAGTCTC 995
 DB 848 rSerThrValThrProSerSerAlaLeuGlyThrThrHsrHsrProPro----- 864
 QY 994 CTGTCCAGACTTAAGCTGAGAGAACTCAGAGATGATGTGCCAGAGCCCGCAC 935

Db 865 -ValProaenThrThraLathrThrhIsgIyArgSerLeuSerProSerProhiSth 884
 QY 934 GCCTGCTGTGACATGTCACAGACAGAGGGCTTGTAGTACATTTCACACATATTG 875
 Db 884 rValcysThrAlaThrThrSerAlaThrSerGlyIleLeuGlyThr-ThhIleIleThc 904
 QY 874 AACACAGTATGGGCTGCCAC----- 852
 Db 904 lProSerThrGlyThrSerHisThrProAlaAlaThrThrGlyThrThrGlnHisSer 924
 QY 852 ----- 852
 Db 924 hrProAlaLeuSerSerProHisProSerSerArgThrThrGlyThrProSerPro 944
 QY 851 --GGAACCTGCCAGAGAGGGGACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 794
 Db 944 lYthrThrThrProGlyHisThrThraLathrSerArgThrThrAlaThrProS 964
 QY 793 CAGCCACAGAGGTTGGAG 734
 Db 964 erLysThrArgThrSerThrLeu-----ProSerG 975
 QY 733 TCCCATAGCTAGAGCTGTAACTACATTTTAACTTGGCATTTATTCACGTTGTCT 674
 Db 975 lnrProThrSerAlaProIleThrThrValValAlaThrMetGlyCysGluProGlnCysAla 995
 QY 673 GGTACATCTGTCTGTCTG--GATGTAGCTTTGCTGTGAGGGT----- 630
 Db 995 rPserGlu-----TriLeuAspTyrSerTyrProMetProGlyProSerGlyVal 1012
 QY 629 -----TCCAGGTAGATGACATGACGTCTGCGAGAGAGAGAGAGAGAGAGAGAG 593
 Db 1012 spHeAspThrTyrSerAsnIleArgAlaIleGlyValAlaValCysGluGlnProLeu 1032
 QY 592 CCTCCACAG 533
 Db 1032 lYleuGlnCysArgAlaGln-Ala-----GlnProGlyValProLeuArgGluLeu 1048
 QY 532 GGTACAG----- 527
 Db 1049 GlyGlnValValGlnCysSerLeuAspPheGlyLeuValCysArgAsnArgGluGlnVal 1068
 QY 526 -----CCAGAGGCTTACATATAGTGTCTTTG----- 497
 Db 1069 GlyLysPheLeuMetCysPheAsnTyrGlnIleArgValPheCysAsnTyrGlyHis 1088
 QY 496 ----- 482
 Db 1089 CysProSerThrProAlaThrSerSerThrAlaThrProSerSerThrProGlyThrThr 1108
 QY 481 TGGCTCATCAGCTTGAAGAAAGACAGCTGCCAGCTGGGGGAGAGAGAGAGAGAG 422
 Db 1109 TrpIleLeuThrGluLeuThrThrAla----- 1118
 QY 421 ACAGAGCTGATGTTCTCTTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 374
 Db 1119 -----ThrThrThrGlyThrGlySerThrAlaThrProThrSerThrLeuArg 1135
 QY 373 -----TGAAGCTTGAAGACAGTATACCTCTCTCATAC 335
 Db 1136 ThrAlaProProProLysValLeuThrThrThraLathrThrProThrValThrSer 1155
 QY 334 CTGAG 275
 Db 1156 LysAlaThrProSer-SerSerProGlyThrAlaThrAlaLeu-----Pr 1170
 QY 274 TGGCATCTCCCTTGAAGAGGCTTCCAGCTGC-----TCACTGTCCAGAGTGT 224
 Db 1170 AlaLeuArgSerThrAlaThrThrProThrAlaThrSerValThrProIleProSer 1190
 QY 223 CTGGTTGTGACCAAG 164
 Db 1190 rSerLeuGlyThrThr-----TprThrArgLeuSerGlnThrThrThrProTh 1206

QY 163 AGCCAGTTGAACACT 148
 Db 1206 rAlaThrMetSerThr 1211

RESULT 8

EDBE11
 Immediate-early protein IE110 - human herpesvirus 1 (strain 17)
 C:Species: human herpesvirus 1
 C:Date: 31-Mar-1988 #sequence_revision 31-Mar-1988 #text_change 17-Mar-2000
 C:Accession: A29152
 J:Gen. Virol. 67, 2365-2380, 1986
 A:Title: Characterization of the IE110 gene of herpes simplex virus type 1.
 A:Reference number: A29152; PMID:87059760; PMID:3023529
 A:Accession: A29152
 A:Molecule type: DNA
 A:Residues: 1-775 <PER>
 A:Cross-references: GB:X04614; NID:G59832; PIDN:CAA28285.1; PID:G59833
 C:Genetics:
 A:Insertions: 19/3; 242/1
 C:Superfamily: herpesvirus immediate-early protein IE110; RING finger homology
 C:Keywords: DNA binding; early protein; transcription regulation; zinc finger
 F:112-162/Domain: RING finger homology <RNG>
 F:116-156/Region: zinc finger C3HC4 motif

Alignment Scores:

Pred. No.:	Score:	Percent Similarity:	Best Local Similarity:	Query Match:	Length:	Matches:	Conservative:	Mismatches:	Indels:	Gaps:
1	117.00	36.00%	31.33%	4.57%	775	47	7	61	35	8

US-09-989-919-15 (1-1397) x EDBE11 (1-775)

QY 1188 CCGGCGCCAGCCCGCCAGGCTGCTGCTCAACACACCTGCTCCAGAGAGAG 1129
 Db 381 ProGlySerAlaProArgProGlyProProAlaSerAlaAlaIleSer---GlyProAla 399
 QY 1128 TTGGCTCATGCTGAGGTG-----TGCACCTTGCCCTGAGAGAGAGAGAGAGAG 1075
 Db 400 ArgProArgAlaAlaValAlaProCysValArgAlaProPro----- 414
 QY 1074 GTGTCCCTCTGTGGAGAGTCAGAGATAGCTCCCGAGTACAGAAATCCACATCTCTG 1015
 Db 415 -----GlyProGlyProArgAlaProAla-ProS 424
 QY 1014 AGCATCCCGCCAGCAAGTCTCTGTCAGACCTTAAGCTGAGAGAACTCCAGAGAGAT 955
 Db 424 yAlaGluProAlaAlaArgProAlaAsp-----AlaArgArgVal----- 437
 QY 954 GTCCAG 895
 Db 438 -----ProGlnSer-HisSerSerLeuAlaGlnAlaAlaAsnGlnGlnSerLeu 455
 QY 894 GCTACATCTTCACATATGTAACACAGTATGG-----CTGCCAGAG 850
 Db 455 sArgAlaArgAlaThrValAlaAlaArgGlySerGlyProGlyValGlnGlyValIle 475
 QY 849 AACCGTCCAG 790
 Db 475 y-----ProSerArgGlyAlaAlaAlaProSerGlyAlaAlaAlaProLeuProSerAlaIle 493
 QY 789 CCACAGGTTGGAG 762
 Db 493 rValGlnGlnGlnAlaAlaValAlaArgPro 502

RESULT 9

G01763
 atrophin-1 - human
 C:Species: Homo sapiens (man)
 C:Date: 21-Dec-1996 #sequence_revision 06-Jun-1997 #text_change 17-Jul-1998

C/Accession: G01763
 R/Margolis, R.L.
 Submitted to the EMBL Data Library, March 1995
 A/Reference number: G08343
 A/Accession: G01763
 A/Status: preliminary; translated from GB/EMBL/DBJ
 A/Molecule type: mRNA
 A/Residues: 1-1184 <MAR>
 A/Cross-references: EMBL:U23851; NID:g915325; PID:g915326
 C/Genetics:
 A/Gene: GDB:DRPLA; B37
 A/Cross-references: GDB:270336; OMIM:125370
 A/Map position: 12p-12p

Alignment Scores:

Pred. No.:	0.209	Length:	1184
Score:	117.00	Matches:	92
Percent Similarity:	32.23%	Conservative:	34
Best Local Similarity:	23.53%	Mismatches:	140
Query Match:	4.57%	Indels:	125
DB:	2	Gaps:	20

US-09-989-919-15 (1-1397) x G01763 (1-1184)

```

QY 1190 GCCCGGCCCCCACC----- 1176
DB 365 AlaProAlaProProMetArgPheProTyrSerSerSerSerSerAlaAlaAla 384
QY 1175 -----CCGCCCAAGTCTGCTCAAAACCA 1149
DB 385 SerSerSerSerSerSerSerSerSerSerAlaSerProPheProAlaSerGlnAlaLeu 404
QY 1148 CTTGCT-----CCCTGACCCAGCTTGGCTCATGCTGAGTGTGACCT 1104
DB 405 ProSerTyrProHisSerPheProProProProHisLeuSerValSerAlaGlnProPro 424
QY 1103 CTGCCCCCTGACCCCTGGGCTGGCTGGAGTGTCCCTGGGGAGGTGACAGATAGCT 1044
DB 425 LysTyrThrGlnProSerLeuProSerGlnAlaValTrpSerGlnGlyProProProPro 444
QY 1043 CCCCAG-----GTACAGATACCCCATCTGGAGCATCCCGCCAG 999
DB 445 ProProTyrGlyAlaGluLeuAlaAsnSerAlaHisProGlyProPheProPro-- 463
QY 998 TCTCCTGTCCAGACTTACAGAGAACTCCAGCATGATGATGCTCCACAGAGCC 939
DB 464 -----SerThrGlyAlaGlnSerThrAlaHisProProValSerThrHisHis 480
QY 938 CACCGCTGTGTGGCATGTGACAGACAGACAGAGGGCTTTAGTACAT----- 888
DB 481 HisHisGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 500
QY 887 -----TTCACACTATTTAACACAGTATGGGCTGCC 855
DB 501 SerGlyProProProProGlyAlaPheProHisProLeuGlu-----GlyGlySerSer 518
QY 854 CACGGAACCTCCACAGGAGGGGACAGCCCCAGTGGAGACTAGCTGAGCTTCTT 795
DB 519 HisHisAlaHisPro--TyrAlaMetSerPro-----SerLeuGlySerLeu 533
QY 794 ACAGCCACACAGGTTGAGAGGAGCATATCCCATGATCTCTGACTAGAGACGACA 725
DB 534 ArgProTyrProProProGlyProAlaHisLeuProPro----- 545
QY 734 GTCCCATAGCTAGGCTGGTAACTACATTTTAACATTTGGCATTTATGGACGTTTGC 675
DB 546 -----ProHis-----SerGlnValSer 551
QY 674 TGGTCATCTGTCTGTCTGGAGATGATAGCTTTGTCTGACAGGGTTCCACAGTGACATGC 615
DB 552 TyrSerGlnAlaGlyPro-----AsnGlyProProValSerSer 564
QY 614 AGTGTGCGACGTCAAGTCACTCCCTCCACAGCAAGCCAGCTGCTGAGAGATATGCAG 555
  
```

```

DB 565 SerSerAlaSerSerSer-SerHisSerGlnGlySerTyrProCys----- 579
QY 554 TGGTCAGCCGAGCCCTCAG-----ATAGTCACGACGAGGCTTCATATATGTGT 504
DB 580 -SerHisProSerProSerGlnGlyProGlnGlyAlaProTyrProPheProProValPr 599
QY 503 CTTGTGTGGATACGAGACCATTTGGTTCATCCAGCTTGGAAAAACGACGCTCCGACC 444
DB 599 othrValThrThrSerSerAlaThrLeuSerThrValIleAlaThrValAlaSerSerPr 619
QY 443 TGGGGGGCAGAGGTGACAGGTAAACAGGGCTGATGTTCTCTTTCTGCA----- 395
DB 619 oAlaGlyTyrIlys-----ThrAlaSerProProGlyProProProTyr 633
QY 394 -GGCAAGACCATTTCTCCTTAATTAGTCCCTTGAAGACAGATACCTCTCTCATCA 336
DB 633 rGlyIlysArg--AlaProSerProGlyAlaTyrIlysThrAlaThrProProGlyTyrIly 652
QY 335 CCTCACACACCCCTCTGACTCCGATGGGGGATCACTATCTATCAACCTCCAGAGAGTC 276
DB 652 sProGlySerProProSerPheArgThrIly-----ThrProPro----- 665
QY 275 CTGCATCTCCCTTGAAGGCTTCTCCACATGCTCACCCTGTCAGGTGCTGTGG 216
DB 666 -----GlyTyrArgGly-----ThrSerProProAlaGlyProGly 677
QY 215 TGACACACAA-----GGCCGACACTGGGC 191
DB 677 YThrPheIysProGlySerProThrValGly 687
  
```

RESULT 10

S50832
 atrophin-1 - human
 C/Species: Homo sapiens (man)
 C/Date: 14-Jul-1995 #sequence_revision 21-Jul-1995 #text_change 07-May-1999
 A/Accession: S50832
 R/Nagafuchi, S.; Yanagisawa, H.; Ohaki, E.; Shitayama, T.; Tadokoro, K.; Inoue, T.; Yan
 Nature Genet. 8, 177-181, 1999
 A/Title: Structure and expression of the gene responsible for the triplet repeat disorder
 A/Reference number: S50832; PMID:95144175; PMID:7842016
 A/Accession: S50832
 A/Status: preliminary
 A/Molecule type: mRNA
 A/Residues: 1-1184 <NAG>
 A/Cross-references: EMBL:D31840
 C/Genetics:
 A/Gene: GDB:DRPLA; B37
 A/Cross-references: GDB:270336; OMIM:125370
 A/Map position: 12p13.31-12p13.3112p-12p

Alignment Scores:

Pred. No.:	0.209	Length:	1184
Score:	117.00	Matches:	92
Percent Similarity:	32.23%	Conservative:	34
Best Local Similarity:	23.53%	Mismatches:	140
Query Match:	4.57%	Indels:	125
DB:	2	Gaps:	20

US-09-989-919-15 (1-1397) x S50832 (1-1184)

```

QY 1190 GCCCGGCCCCCACC----- 1176
DB 366 AlaProAlaProProMetArgPheProTyrSerSerSerSerSerAlaAlaAla 385
QY 1175 -----CCGCCCAAGTCTGCTCAAAACCA 1149
DB 386 SerSerSerSerSerSerSerSerSerSerAlaSerProPheProAlaSerGlnAlaLeu 405
QY 1148 CTTGCT-----CCCTGACCCAGCTTGGCTCATGCTGAGTGTGACCT 1104
DB 406 ProSerTyrProHisSerPheProProProThrSerLeuSerValSerAlaGlnProPro 425
  
```


OY	1103	CTGCCCCGACCCCTGGGCTGGCCCTGGAGAGTCTCCCTGGGAGGTCAGAGATAGGCT	1044
Db	426	LysTyrThrGlnProSerLeuProSerGlnAlaValTyrPheGlnGlyProProProPro	445
OY	1043	CCCCAG-----GTACAGATACACCCACATCTGTGAGATATCCCCGCCAAG	999
Db	446	ProProTyrGlyArgLeuLeuAlaMetSerAlaIleHisPheGlyProPheProPro---	464
OY	998	TCTCTGTCCAGACTTAAGCTTGAGAAACATCTCCACGATGGATTCTCCCAAGCAGCCCC	939
Db	465	-----SerThrGlyAlaGlnSerThrAlaHisProProValSerThrHisHisHis	481
OY	938	CACCGCCTGTGTGGCATGTGACGAGAGACGAGAGCGGGGCTTTAGCTACAT-----	888
Db	482	HisHisGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnHisHisGlyLeu	501
OY	887	-----TCTACACTATTGAACACAGTGAATGGGGCTGCC	855
Db	502	SerGlyProProProProGlyAlaPheProHisProLeuGlu-----GlyGlySerSer	519
OY	854	CACGGAACCGTCCAGAGGGAGGCGCACACCCCACTGGGAGAGACTAGCTGACGTTGTT	795
Db	520	HisHisAlaHisPro---TyrAlaMetSerPro-----SerLeuGlySerLeu	534
OY	794	ACAGCCCAACAGGGTTGAGAGAGGACGACTACATCCCCCATGATTCCTGACATAGAGCCACA	735
Db	535	ArgProTyrProProGlyProAlaHisLeuProPro-----	546
OY	734	GTCCCATAGCTAGGCTGTGTAACCTACATTTTAACATTGGCATTATTGCACGTTTGCC	675
Db	547	-----ProHis-----SerGlnValSer	552
OY	674	TGGTCACATCTGTCTGTGGGATGTAGCTTGTCTGAGGGGTTCCAGGTGACATGAC	615
Db	553	TyrSerGlnAlaGlyPro-----AsnGlyProProAlaSerSer	565
OY	614	AGTGTGGCAGTGACAGTCACTCCCTCCACAGGACCCAGCCTGAGGATGATGACG	555
Db	566	SerSerAenSerSerSerSerThrSerGlnGlySerTyrProCys-----	580
OY	554	TGGTCAGCGGACCCCTCAG-----ATAGTCAGCCACAGAGGCTTCACATATGTGT	504
Db	581	SerHisProSerProSerGlnGlyProGlnGlyAlaProTyrProPheProProValPr	600
OY	503	CTTGTGGATGACGAGGACCACTGTGCTCCATCAGCTCTTGAAAGAACAGCAGTGCCAGC	444
Db	600	othrValThrThrSerSerAlaThrLeuSerThrValIleAlaThrValAlaSerSerPr	620
OY	443	TGGGGGACAGAGGTGAGAGTAAACAGGCGCTGAGTTCCTCTTCTGCA-----	395
Db	620	AlaGlyTyrHis-----ThrAlaSerProProGlyProProProTyr	634
OY	394	GGCAGAGCAATTCCTCCTAATTGACATGCGCTTGAAGACAGTAACTCTCTCATCA	336
Db	634	rgIlyLysArg---AlaProSerProGlyAlaTyrTrpThrAlaThrProProGlyTyr	653
OY	335	CCTCAGACACCCCTGTGACTTCGATGGGGGATCATCTATCATCACTCCAGAGAGGTC	276
Db	653	sProGlySerProProSerPheArgThrGly-----ThrProPro-----	666
OY	275	CTGCCATCTCCCTTGGAAAGGCTTCTCCACATGCGTCACTGTGCCAGTGTGTTGGTGG	216
Db	667	-----GlyTyrArgGly-----ThrSerProProAlaGlyProGly	678
OY	215	TGACCACAAA-----GGCCGACACGTGGC	191
Db	678	YThrPheLysProGlySerProIleHisValGly	688

RESULT 11
T46500
hypochemical protein DKFZp434D098.1 - human (fragment)
C|Species: Homo sapiens (man)
C|Date: 04-Feb-2000 #sequence_revision 04-Feb-2000 #next_change 04-Feb-2000

C:Accession: T46500
R:Ottengaelder, B.; Obermaier, B.; Newes, H. W.; Gassenhuber, J.; Wiemann, S
submitted to the Protein Sequence Database, January 2000
A:Reference number: Z25031
A:Accession: T46500
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-649 <AAA>
A:Cross-references: EMBL:AL137564
A:Experimental source: adult testis, clone DKFZp434D098
C:Genetics:
A>Note: DKFZp434D098.1

```

Alignment Scores:
Pred. No.:      0.304      Length:      649
Score:          115.00     Matches:     95
Percent Similarity: 33.08%  Conservative: 34
Best Local Similarity: 24.16%  Mismatches: 103
Query Match:    4.43%      Gaps:        158
DB:              2
US-03-989-919-15 (1-1397) x T46500 (1-649)

```

QY	1191	GGCCCCGGGCCCCACACCC-----GCCCAAGATCTGGCTCAAAACACACCTGCTCCCTAGC	1139
Db	205	GlySerGlyProGluProIleuAlaProSerPro-----ValSer	217
QY	1134	CCAGCTCTTGAGCTCACTGAGAGGTGACACTCTGACCCCTGACCCCTGGGCTGGCTGGGA	1075
Db	218	ProThrPhePro-----ProSerSerProSerAspTrpPro--GlnGlu	231
QY	1074	GTCGTCCTCTGGGAGGTCAAGATAGACCTCCCAAGGTACAGAAATACACCACTAC--	1018
Db	232	ArgSerProGlyGlyHisSerAspGlyAlaSerProArgSerProValProThrThreIleu	251
QY	1017	-----TGGAGCATCCCGCCGAAGCTCTCGTCCAGACCTAAGC	979
Db	252	ProGlyIleuAlaGHisAlaProTrpGlnGlyProArgGly-----ProProAspSer	268
QY	978	TGAAGAAACTCA-----CGATGATTTGTCCTCCACGACG	943
Db	269	ProAspGlySerProIleuThrProValIleuSerGlnMetProTrpIleuValAlaSerPro	288
QY	942	CCCCCACC-----CTGCTGTGGCATGTGAGCAGCAGCAGCAGAGGGGT	898
Db	289	GluProProGlnSerSerProThrProAlaPheProIleuAlaAlaSer-----	304
QY	897	TTAGCTACATCTCACTATTGAAACAGATGAGGGCTGCCCA-----	853
Db	305	-----TyrAspThrAsnGlyIleuSerGlnProIleuProGluIys	318
QY	852	CGAACCCTCCCAAGGAGGGCAGACACCC--CAGTGGGAGACTAGCCTGAGCTTGCT	796
Db	319	ArgHisIleuProGlyProGlyGlnGlnProIleuProTrpGly-----ProGlnIleuAla	336
QY	795	TACAGCCCAAGGGTGGAGAGGAGCAGATACCCCATGATCTCTGACTGAGACGACG	736
Db	337	SerSerProAla-----ArgGly-----	342
QY	735	AGTCCCATAGTAGCTGGTAAACTCACTATTTAACATTTGGCATTTAGACGTTTGC	676
Db	343	-----IleSerHisHisValThrPheAla	350
QY	675	CTGGTCACATCTGTCTGTCTGGAGATGTTAGCTTTGTCTGCAGGGGTTCCCAAGTCACATG	616
Db	351	ProIleuLeuSer-----	354
QY	615	CAGTGTGCCAGTGCAGATCACTCCCTCCACAGGCAAGCCAGAGCTGTGAGATATGACG	556
Db	355	-----AspAsnValProGlnThrProGluProProThrGlnIleuSerGln	369
QY	555	CTGGTCAGCCGAGCCTTCAGATAGTACGCCAGAGGCTTCACATATGTGGTCTTTGTTGG	496

Db 370 SerAsnValIysPheValGlnAspThrSerLys-----PheTyr 382
 QY 495 GATCAGGAGCACTTGCTGCATCC-----AGTCTGAAAAAGACAGCTGCCGAC 445
 Db 383 -TyrLysProHisLeuSerArgAspGlnAlaIleAlaLeuLysAspPro-- 401
 QY 444 CTGGGGGGCGAGGTAGAGGTAAACAGGGGTGTTCTCTTCTGCGAGCAAGCA 385
 Db 402 -----GlyAlaPheLeuIle--ArgAspSerHl 410
 QY 384 TTCTCCCTAATTGACTGCTTGAGAGACAGGTATACCTCTCTCATCAGCAGACC 325
 Db 410 sSer-----PheGlnGlyAlaTyr-GlyLeuAlaLeuLysValAlaThr 425
 QY 324 CCTCTGACTTCCAGTGGGGGATCATATCTATCACTCCAGAGAGGTCTGCCATCTCC 265
 Db 425 ro-----ProPheSerAlaGln-----P 431
 QY 264 CTGGAAGAGGC-----TTCTCCCACTGGCTC-----ACCTGTCGAC 229
 Db 431 roTrrpLysGlyAspProValGlnGlnLeuValArgHisPheLeuIleGlnThrGlyProl 451
 QY 228 GTGGT-----CTGGTGTGACCAACAAAGCCCGACACATGGGC-----ATGGCTCTCA 181
 Db 451 ysglyValLysIleLysGlyCysProSerGluProTyrPheGlySerLeuSerAlaLeu 471
 QY 180 CAGACATCCACAGCTTACGCCAGTT 155
 Db 471 alserGlnHisSerIleSerProIle 479
 RESULT 12
 A33380
 interleukin-4 receptor precursor - mouse
 N/Alternate names: IL-4 receptor
 C/Species: Mus musculus (house mouse)
 C/Date: 30-Jun-1993 #sequence revision 30-Jun-1993 #text_change 01-Dec-2000
 C/Accession: A33380; B33380; G33380; A34861; I54232
 R/Mosley, B.; Beckmann, M.P.; March, C.U.; Idzerda, R.L.; Gimpel, S.D.; Vandenbos, T.; H
 Wimer, M.B.; Cosman, D.; Park, L.S.
 Cell 59, 335-348, 1989
 A/Title: The murine interleukin-4 receptor: molecular cloning and characterization of se
 A/Reference number: A90911; MUID:90030408; PMID:2805066
 A/Molecule type: mRNA
 A/Accession: A33380
 A/Residues: 1-810 <MO1>
 A/Cross-references: GB:W27959; NID:G198363; PIDN:AAA39299.1; PID:G309408
 A/Accession: B33380
 A/Molecule type: mRNA
 A/Residues: 1-258 <MO2>
 A/Accession: C33380
 A/Molecule type: mRNA
 A/Residues: 1-224, PSNENL, <MO3>
 A/Cross-references: GB:W27960; NID:G198365; PIDN:AAA39300.1; PID:G309409
 A/Note: part of this sequence, including the amino end of the mature protein, was confir
 A/Note: three forms of cDNA were isolated; the longest encodes extracellular, transmembr
 acellular domain and may encode a soluble form of the receptor
 R/Harada, N.; Castle, B.E.; Gotman, D.M.; Itoh, N.; Schreurs, J.; Barrett, R.L.; Howard,
 Proc. Natl. Acad. Sci. U.S.A. 87, 857-861, 1990
 A/Title: Expression cloning of a cDNA encoding the murine interleukin 4 receptor based c
 A/Reference number: A34861; MUID:90138976; PMID:2405398
 A/Accession: A34861
 A/Molecule type: mRNA
 A/Residues: 1-810 <HAR>
 A/Cross-references: GB:W29654; NID:G198346; PIDN:AAA39297.1; PID:G309407
 R/Wrighton, N.C.; Campbell, L.A.; Lee, F.D.
 Growth Factors 6, 103-118, 1992
 A/Title: The murine interleukin-4 receptor gene: Genomic structure, expression and poten
 A/Reference number: I54232; MUID:92265333; PMID:1534014
 A/Accession: I54232
 A/Status: Preliminary; translated from GB/EMBL/DBJ
 A/Molecule type: DNA
 A/Residues: 1-73, 'I', 75-333, 'P', 335-810 <RES>
 A/Cross-references: GB:W64879; NID:G198359; PIDN:AAB59727.1; PID:G198361

C/Genetics:
 A/Intons: 24/1; 71/2; 122/1; 172/3; 225/1; 258/2; 284/3; 301/2
 C/Superfamily: Interleukin-4 receptor; cytokine receptor homology
 C/Keywords: alternative splicing; cytokine receptor; glycoprotein; transmembrane protein
 F/1-25/Domain: signal sequence #status predicted <SIS>
 F/26-810/Product: interleukin-4 receptor #status experimental <MAT>
 F/26-233/Domain: extracellular #status predicted <EXT>
 F/24-219/Domain: cytokine receptor homology <CRS>
 F/234-257/Domain: transmembrane #status predicted <TM>
 F/258-810/Domain: intracellular #status predicted <INT>
 F/12,129,135,163,177/Binding site: carbohydrate (Asn) (covalent) #status predicted

Alignment Scores:
 Pred. No.: 0.366 Length: 810
 Score: 114.00 Matches: 102
 Percent Similarity: 32.71% Conservative: 38
 Best Local Similarity: 23.83% Mismatches: 123
 Query Match: 4.45% Indels: 166
 DB: 1 Gaps: 26

US-09-989-919-15 (1-1397) X A33380 (1-810)

QY 1133 CAGCTTGGCTGATGATGAGTGACCTTGCCCTGACCCCTGG-----CTG 1083
 Db 428 GlnSerAlaLeuAlaGlnSerCysSerProLeuProSerGlySerGlyGlnAlaSerVal 447
 QY 1082 GCCTGGAGTGT---CCCTGTCGGAGGTGACAGATAGCTCCCGAGTACAGATCAG 1026
 Db 448 SerTrrpAlaCysLeuProMetGlyProSerGlnGlnAlaThrCysGlnValThrGlnGln 467
 QY 1025 CCA---CATCTGGA----- 1014
 Db 468 ProSerHisProGlyProLeuSerGlySerProAlaGlnSerAlaProThrLeuAlaCys 487
 QY 1013 -----GCA 1011
 Db 488 ThrGlnValProLeuValLeuAlaAspSerProAlaTyrArgSerPheSerAspCys 507
 QY 1010 TCCCGCCCAAGTCTCTGTCAGACCTTAGTGAGAACACTCCACATGATGG--- 954
 Db 508 SerProAlaProAspProGlyGlnLeuAlaProGlnGlnGlnAlaAlaAspHisLeuGln 527
 QY 953 -----TCCCGACGAGCCCGCCACCGCTG-----CTGGGACATGG 918
 Db 528 GlnGlnGlnProProSerProAlaAspProHisSerSerGlyProProMetGlnProVal 547
 QY 917 CAGCAGCAGCAGCAGGGGCTTAGCTTACATCTTCACATTAATTAACACAGTATGGGCT 858
 Db 548 GlnSerTrrpGlnGln-----IleLeuHisMetSerValLeuGlnHis-----GlyAla 563
 QY 857 GCCCAGCAGAACCTGCCAGG----- 837
 Db 564 AlaAlaGlySerThrProAlaProAlaGlyGlyTyrGlnGlnPheValGlnAlaValLys 583
 QY 836 -----AGGGCAGACAGCCCGCAGTGGGAGACTAGCTG 804
 Db 584 GlnGlyAlaAlaGlnAspProGlyValProGlyValAlaGlnProSerGlyAspProGlyTyr 603
 QY 803 -----AGCTTGCTTACAGCCACAGGTT----- 780
 Db 604 LysAlaPheSerSerLeuSerSerSerSerAlaGlyIleArgGlyAspThrAlaAlaGly 623
 QY 779 -----GAGAGGAGACTCATCCCATATCTCTGAGTACAGCA----- 738
 Db 624 ThrAspAspGlyHisGlyGlyTyrLysPro---PheGlnAspProValProAsnGlnSer 642
 QY 737 -----GCAGTCCCATAGCTAGAGCTGTAACTCATTTTAACTTGGCATTTATGGA 684
 Db 643 ProSerSerValPro-----LeuPheThrPheGlyLeuAspThr 655
 QY 683 CGTTGCTCTGCTGATCATCTGTC-----TGTCTGGA 652
 Db 656 GlnLeuSerProSerProLeuAsnSerAspProProlLysSerProProGlnCysLeuGly 675

QY 544 -AGCCCTCAGATAGTCAGCCAGAGCCCTTACATATGTGCTTTGGATCAGGAC 486
 Db 274 pATGATleatpPhc1yAlaProGlnAlaPheLeuAlaGlyLeuLeuAlaThrVal 294
 QY 485 CACTTGGCTTCATCCAGTCTTGAAMAGACCACTGCCAGCTGGGGGAGAGGTGAG 426
 Db 294 a-ValGlyThrAlaArgAlaGlyLeuGlnProAlaAspMet----- 308
 QY 425 GGTACAGAGGCTGATGTTTCTCTTTGTGAGGCAAGACATTCCTTAATTGATGC 366
 Db 308 ----- 308
 QY 365 CTGAAGACAGATATACCTCTCTATCATCACTCAGACCCCTTGATCCGATGGG 306
 Db 309 -----AlaAlaProProThr-LeuProGlnPro 317
 QY 305 GATCACTATCTATCAACCTCCAGAGAGTCCCTGATCTCCCTTGAGAGGCTTCCCA 246
 Db 318 -----Pro 318
 QY 245 CTGGCTCAGCTGTCCAGGTGCTGTGTGTGATACCAAGGCCCACTGGGATGC 186
 Db 319 ArgAlaHisGlyGlnHisGlyGlnHisHisGlnLeuProPheLeuGlyHisAsp 338
 QY 185 TCTCAGACATCCAGAGCCTCAGCCAGGTTGAACACTGAAGAGGAGGAGCTCCCGCT 126
 Db 339 GlnHisHis-----GlyGlnHisLeuArgVal 347
 QY 125 GGTAGATGAGCCAGAGCGTAGT-----CTTCTGGGGATGG 87
 Db 348 GlyGlnHisGlyArgAlaHisAspValLeuProGlyHisLeuGlnGlyTyr 367
 QY 86 TGCTGTCTGGGA-----TACACTGTAAGTGGTACTGC 54
 Db 368 GlyCysTyrAsnLeuSerAspTyrHisGlnGlyThrHisValCys 382

RESULT 14

T09059
 notch4 - mouse
 C/Species: Mus musculus (house mouse)
 C/Date: 11-Jun-1999 #sequence_revision 11-Jun-1999 #text_change 11-Jan-2000
 C/Accession: T09059
 R/Romen, L.; Mahliras, G.; Qin, S.; Ahearn, M.E.; Dankers, C.; Loretz, C.; Sc
 submitted to the EMBL Data Library, October 1997
 A/Description: Sequence of the mouse major histocompatibility locus class III region.
 A/Reference number: Z16543
 A/Accession: T09059
 A/Status: preliminary; translated from GB/EMBL/DBJ
 A/Residues: 1-1964 <ROW>
 A/Cross-references: EMBL:AF030001; NID:G2564945; PID:G2564947
 A/Molecule type: DNA
 A/Genetics:
 A/Gene: notch4
 A/Map position: 17
 A/Intons: 22/1; 49/2; 148/1; 264/1; 305/1; 384/1; 436/1; 501/1; 539/1; 577/1; 618/1; 67
 1679/3; 1129/1; 1761/3
 C/Superfamily: unassigned EGF-related proteins; EGF homology
 C/Keywords: receptor; signal transduction
 F/514-545/Domain: EGF homology <EGF>

Alignment Scores:

Prod. No.: 0.365
 Score: 114.00
 Percent Similarity: 28.84%
 Best Local Similarity: 22.11%
 Query Match: 4.51%
 DB: 2
 Gaps: 37

US-09-989-919-15 (1-1397) x-T09059 (1-1964)
 QY 6 TGGACCTGTACCGAGCGGCGAGATATCTGCAGAACTCCAGCG-----CAAGCAGCAGTA 59
 Db 6 TGGACCTGTACCGAGCGGCGAGATATCTGCAGAACTCCAGCG-----CAAGCAGCAGTA 59

Db 903 CysIleAspThrGlySerSerTyrPheCysArgCysProProGlyPheGlnGlyLeu 922
 QY 60 CCGAGTACAGAGTATCCAGACAGACAGACATCCCGAG----- 97
 Db 923 CysGlnAspAsnValAsnProCysGlnProAsnProCysHisHisGlySerThrCysVal 942
 QY 98 -----GAAAGCTAC-----CGTCTGGGCCCATCTTACACACCGGAGCTGC 142
 Db 943 ProGlnProSerGlyTyrValCysGlnCysAlaProGlyTyrGlnGlyGlnAsnCys 961
 QY 143 CTTTACGTGTTAACTGGCTGAGCTGATGATGCTGTGAAGACCATCCAGTCCG 202
 Db 962 -----SerIleValLeuAspAlaCysGlnSerGln--ProCysHis 974
 QY 203 GCCTTTGTGTGACCAACAGACACCTTGAAGAGGAGCCAGTGGAGAA-----GCC 256
 Db 975 -----AsnHisGlyThrCysThrSerArgProGlyGlyPheHisCysAla 989
 QY 257 CTTCCAGAGGAGATGGACAGCTCTCTGAGAGTTG-----ATAGATAGT 301
 Db 990 CysProProGlyPheValGly-----LeuArgCysGlnGlyAspValAspGlyCysLeu 1007
 QY 302 GATCCCCATCGAAGTCAAGAGGGGTGCT-----GAGGTATGAGAGAGAGTAT 352
 Db 1008 AspArgProCysHisProSerGlyThrAlaAlaCysHisSerLeuAlaAsnAlaPheTyr 1027
 QY 353 -----ACGTGCTTCAGAGC----- 367
 Db 1028 CysGlnCysLeuProGlyHisThrGlyGlnArgCysGlnValGlnMetAspLeuCysGln 1047
 QY 368 AGTCAATATGAGGAGAGATGCTTGTCTTCCAGAGAGAAATCAGACCT-----GT 421
 Db 1048 SerGlnProCysSerAsnGlyGlySerCysGlnIleThrGlyProProProGlyHis 1067
 QY 422 TACCTTACCTCTGCCCCCAGGTGGAG----- 452
 Db 1067 ThrHisHis-----CysProGlyGlyPheGlnGlyProThrCysSerHisLeuAlaLeu 1086
 QY 453 -----CTGCTCTTTTCAAGACTGATGAGGCAAGTGGTCCGTATCCCAAGACAC 508
 Db 1086 rCysGlyIleHisHisCysHisAsnGlyGlyLeuCysLeuProSerPro----- 1102
 QY 509 ATATGTAAGGCGCTGTGCTGACT-----ATCTGAGGCTTGGCTGACA-- 554
 Db 1103 -----LysProGlySerProProLeuCysAlaCysLeuSerGlyPheGlyGlyProAs 1120
 QY 555 -----GCTGACTATCTCAGCAGGTGGGCTTGCCT-----GTGAGGAGCT 595
 Db 1120 rCysLeuThrProProAlaProProGlyCysGlyProProSerProCysLeuHisAsnGly 1140
 QY 596 GACTTGCACTGGCAGACATGCATGCACCTGGGAACCCCTGCAGACAAAGCTAACATCC 655
 Db 1140 yThrCysThrGlnThrProGly-----LeuGlyAsnProGlyPheGlnCysThrCysPr 1158
 QY 656 AGACAGACAGATGTGACAGACAAAGCTGCAAT----- 689
 Db 1158 oProAspSer-----ProGlyProArgCysGlnArgProGlyAlaSerGlyCysGlnGly 1176
 QY 690 -----AATGCCAATGTTAAATGATGAGTTTACCAAGCTAGCT 727
 Db 1176 yArgGlyGlyAspGlyThrCysAspAlaGlyCys-----SerGlyProGlyGlyAs 1193
 QY 728 ATGGAGC----- 734
 Db 1193 pThrAspGlyGlyAspCysSerLeuGlyValProAspProTyrGlyCysProProHis 1213
 QY 735 -----TCTGAGCTTCTGACGAGATCATG----- 761
 Db 1213 sSerGlnCysThrPheLeuPheArgAspGlyArgCysHisProGlnCysAspSerGlyGly 1233
 QY 762 -----GGGTATAGTGCCTCTCCACCTGTG----- 788
 Db 1233 uCysLeuPheAspGlyTyrAspCys-----GlnIleProLeuThrCysIleProAlaTyrAs 1252

```

QY 789 -----GGCTGTAAAGCAAGCTCA 805
Db 1252 pGlnTyrCyArGAspHisPheHisAenGlyHisCysGluLysGlyCysAenAenAlaG 1272
QY 806 GGGTACTGTCCCACTGGGGGCTGTGCTCCCTCCCTGGAGAGGTTCCGTGGGACGCCCAT 865
Db 1272 uCysGlyTTPAspGlyGlyAspCysArgProGluGluAspSerGluGlyArgProSe 1292
QY 866 CACTGTGTCAATAGTGTGAATGTAGCTAAAGCCCTGTGTGTGTGTGTGTGTGTGTGTGT 925
Db 1292 rLeuAla-----LeuLeuValValLeuArgProPr 1302
QY 926 ACAGCAGCGCGGTGGGCTGCTGGGACAATCCATGCTGAGTGTGTGTGTGTGTGTGTGTGT 985
Db 1302 o-----AlaLeuAspGlnGlnLeuLeuAlaLeuAlaArgValLeuSe 1316
QY 986 TCTGACAGAGACTGTGGCGGGGAGTCTCCAGAGATGGGTATCTGTACTGGGAG 1045
Db 1316 rLeuThrLeuArgValGly-----LeuTrpVal-----Ar 1326
QY 1046 GCATCTCTGACTCCCGACAGGGGACACTCCA----- 1079
Db 1326 gLysAspSerGluGlyArgAsnMetValPheProTyrProGlyThrArgAlaLysGluG 1346
QY 1080 -----GGCCAGCCCGAGGAGTCCAGGGGACAGAGTGCACACTCAGCATGAGCCA-- 1127
Db 1346 uLeuSerGlyAlaArgAspSerSerSerTrpGluArgGlnAlaProProthrGlnProle 1366
QY 1128 -----AGACTGGGCTCAGGG-----AG 1144
Db 1366 uGlyLysGluThrGluSerLeuGlyAlaGlyPheValValMetGlyValAspLeuSe 1386
QY 1145 CAGGTGTGTTTGTAGCCAGACCTGGGGGGG-----GGGTGGGGCGGGGCTTCTGTGCTC 1201
Db 1386 rArgCysGlyProGluHisProAlaSerArgCysProTrpAspSerGlyLeuLeuLeuAr 1406
QY 1202 ATTTGCTTTCATGAAGACCTCAAGAGCCAAAGCCAGGCTTTCCTTCTC 1256
Db 1406 gPheLeu-----AlaAlaMetAlaAlaValGlyAlaLeuGlnProLeuLeu 1421

RESULT 15
A48060
erythroid Kruppel-like factor EKLf - mouse
C:Species: Mus musculus (house mouse)
C:Date: 21-Jan-1994 #sequence_revision 18-Nov-1994 #text_change 05-Nov-1999
C:Accession: A48060
R:Miller, I.U.; Bieker, J.J.
Mol. Cell. Biol. 13, 2776-2786, 1993
A:Title: A novel, erythroid cell-specific murine transcription factor that binds to the
A:Reference number: A48060; MUID:93233640; PMID:7682653
A:Accession: A48060
A:Status: Preliminary
A:Molecule type: nucleic acid
A:Residues: 1-376 <ML>
A:Cross-references: GB:W92700; NID:9193011; PID:AAA37546.1; PID:9193012
A>Note: sequence extracted from NCBI backbone (NCBIN:129821, NCBI:P.129822)

Alignment Scores:
Pred. No.: 0.487 Length: 376
Score: 112.50 Matches: 80
Percent Similarity: 32.23% Conservative: 27
Best Local Similarity: 24.10% Mismatches: 97
Query Match: 4.39% Indels: 128
DB: 2 Gaps: 18

US-09-989-919-15 (1-1397) x A48060 (1-376)
QY 1179 ACCCGCCGCGCCAGT-----CCMGCTCAAAACACACACTGTCTCCGTGACCCCACTC 1129
Db 98 ThrAsnPheProGlySerGluSerProGlyThrSerArgThrCysAlaLeuAlaProSer 117
QY 1128 TTGGCTCATGCTGAGGTGTGACACTTGCCCTGACCCCTGAGCTGGGCTGGAGATGTCC 1069

```

```

Db 118 ValGlyPro-----ValAlaGlnPheGluProProGluSerLeuGlyAlaTyr 133
QY 1068 CCTGTGGAGAGGTCAAGATAGCTCCCA-----GGTACAGAAATACCCCATCTCTGAGC 1012
Db 134 AlaGlyGlyProGlyLeuValThrGlyProLeuGlySerGluGlnHisThrSerTrpAla 153
QY 1011 ATCCCGCCGCAAGTCTCTGTCCAGACTTAAGCTGAGAGAACACTCCAGATGATGTTC 952
Db 154 HisPro-----ThrProGln----- 158
QY 951 CCCAGCAGCCCCCACCCTGCTGTGATGTGCAGCAGCAGCAGAGGCGCTTAACT 892
Db 159 -----ProProAlaProGluProPheValAlaProAlaLeuAlaProGlyLeuAla 175
QY 891 ACATTCTACACTATTGAACACAGATAGGGGCTGCCACGGAACCGTCCCAAG----- 838
Db 176 -----ProLysAlaGlnProSerTyrSer 183
QY 837 -----GAGGAGCAGAGCCCCCAGTGGGGA----- 814
Db 184 AspSerArgAlaGlySerValGlyGlyPhePheProArgAlaGlyLeuAlaValProAla 203
QY 813 -----GACTACCTGAGCTTCAACGCCCAAG 784
Db 204 AlaProGlyAlaProTyrGlyLeuLeuSerGlyTyrProAlaLeuTyrProAlaProGln 223
QY 783 GGTGGAGAGAGCATATACCCCATGATTTCTGAGCTAGAGCAGCAGCAGTCCCATAGCT 724
Db 224 ---TyrGlnGly---HisPheGlnLeuPheArgGlyLeuAlaAlaProSerAlaGlyGly 241
QY 723 AGCGTGTAACTACATATTTAATTTGATTTGATTTGATTTGATTTGATTTGATTTGATTTG 664
Db 242 ThrAlaProProSerPheLeuAsn----- 249
QY 663 GTCTGTCTGGAGATTTAGCTTTGTCTGACAGGGTTCCAGGTGACATGACAGTCTGCCG 604
Db 250 ---CysLeuGly---ProGlyLysThrValAlaThrGluLe 260
QY 603 TGCAGTCACTCCCTCCACAGCAGCAGCCAGCTGTCTGAGATGATGATGATGATGATGATG 544
Db 260 uGlyAlaThrAlaAlaAlaGlyAspAlaGlyLeu-----SerProGlyThrAl 276
QY 543 GCCCTGAGATAGGTGAGCCAGAGGCTT-----CACATATGTGG 505
Db 276 aProProLysArgSerArgArgThrLeuAlaProLysArgGlnAlaAlaHisThrCysG 296
QY 504 TCTTGTGGG-----ATCAGGAGCACTTGCTCC 475
Db 296 yHisGluGlyCysGlyLysSerTyrSerLysSerHisLeuLysAlaHisLeuArgTh 316
QY 474 ATCCAGTCTTGAAGAAAGACAGCTGCCGACCTG---GGGCGAGAGCTAGA----- 426
Db 316 rHisThrGlyGlyLysProTyrAlaCysSerTrpAspGlyCysAspTrpArgPheAlaAr 336
QY 425 -----GGTACAGGAGCTGAGTTTC----- 405
Db 336 gSerArgGluLeuThrArgHisTyrArgLysHisThrGlyHisArgProPheCysCysG 356
QY 404 -----TCTTCTGAGCAGCAACCAT 384
Db 356 yLeuCysProArgAlaPheSerArgSerArgHis 367

Search completed: June 24, 2003, 13:59:53
Job time : 95 secs

```


GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on: June 24, 2003, 13:47:02 ; Search time 119 Seconds
(without alignments)
4837.778 Million cell updates/sec

Title: US-09-989-919-15
Perfect score: 2527
Sequence: 1 ggtgcgcgcctcaccgga.....aaaaaaaaaaagcggtc 1397

Scoring table: BLOSUM62
Xgapop 10.0, Xgapext 0.5
Ygapop 10.0, Ygapext 0.5
Fgapop 6.0, Fgapext 7.0
Delop 6.0, Delext 7.0

Searched: 671580 segs, 206047115 residues
Total number of hits satisfying chosen parameters: 1343160

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame+ n2p model -DRV=xlp
-Q/cgnt2.1/USFTO.spool/US09989919/runat.24062003.102047.19184/app.query.fasta.1.1543
-DB=SPREMBL_21 -QFMT=fasten -SUFFIX=rspt -MINMATCH=0.1 -LOOPEC=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45
-DOCAIIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=pcr -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000
-USER=US09989919 @CGN 1.158 @runat.24062003.102047.19184 -NCPU=6 -ICPU=3
-NO_MAP -LANG=OTHER -NEG_SCORES=0 -WAIT -DSPLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREDS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :
1: SPREMBL_21:*
2: sp_archaea:*
3: sp_bacteria:*
4: sp_fungi:*
5: sp_human:*
6: sp_invertebrate:*
7: sp_mammal:*
8: sp_mhc:*
9: sp_organelle:*
10: sp_phage:*
11: sp_plant:*
12: sp_rodent:*
13: sp_virus:*
14: sp_vertebrate:*
15: sp_unclassified:*
16: sp_virius:*
17: sp_bacteriophage:*
18: sp_bacteriophage:

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

Result No.	Score	Query Match	Length	DB ID	Description
1	519.5	20.6	207	4 Q96109	Q96109 homo sapien

2	431	17.1	268	11 Q8VBZ6	Q8VBZ6 mus musculus
C 3	143	5.6	1394	4 Q00512	Q00512 homo sapien
C 4	135.5	5.3	631	6 Q9N1P1	Q9N1P1 bos taurus
C 5	134	5.2	4123	4 Q75851	Q75851 homo sapien
C 6	127.5	5.0	4957	4 Q14687	Q14687 homo sapien
C 7	127.5	5.0	5262	4 Q14686	Q14686 homo sapien
C 8	124.5	4.9	818	6 Q9N1P0	Q9N1P0 bos taurus
C 9	124.5	4.9	1280	11 Q9EPX2	Q9EPX2 mus musculus
C 10	123.5	4.9	669	4 Q75441	Q75441 homo sapien
C 11	122.5	4.8	647	10 Q2ARH1	Q2ARH1 brassica na
C 12	121.5	4.7	1400	4 Q9UP57	Q9UP57 homo sapien
C 13	120.5	4.7	763	4 Q9UPX0	Q9UPX0 homo sapien
C 14	120.5	4.7	1236	4 Q15398	Q15398 homo sapien
C 15	120	4.7	815	5 Q9V9W8	Q9V9W8 drosophila
C 16	119	4.6	457	5 Q961D0	Q961D0 drosophila
C 17	119	4.6	749	5 Q9W384	Q9W384 drosophila
C 18	118	4.7	704	11 Q9CWS0	Q9CWS0 mus musculus
C 19	118	4.6	1063	12 Q40956	Q40956 rubella vir
C 20	118	4.6	1063	12 Q8VA12	Q8VA12 rubella vir
C 21	118	4.6	1063	12 Q8VA10	Q8VA10 rubella vir
C 22	118	4.6	3570	4 Q99552	Q99552 homo sapien
C 23	117.5	4.6	596	11 Q9CVP3	Q9CVP3 mus musculus
C 24	117.5	4.6	749	5 Q9B1W6	Q9B1W6 drosophila
C 25	117	4.6	995	4 Q9Y2W4	Q9Y2W4 homo sapien
C 26	117	4.6	1182	4 Q94995	Q94995 homo sapien
C 27	116.5	4.6	970	12 Q8QZ06	Q8QZ06 rana tigris
C 28	116.5	4.6	1283	5 Q8T5H0	Q8T5H0 anopheles g
C 29	116.5	4.6	1511	4 Q75412	Q75412 homo sapien
C 30	116.5	4.6	1566	4 Q9P2R6	Q9P2R6 homo sapien
C 31	116	4.5	726	11 Q924S6	Q924S6 mus musculus
C 32	116	4.5	957	4 Q9UKN0	Q9UKN0 homo sapien
C 33	115.5	4.6	840	4 Q9Y2K5	Q9Y2K5 homo sapien
C 34	115	4.5	394	12 Q9J2I0	Q9J2I0 macaca mla
C 35	115	4.5	649	4 Q9NT29	Q9NT29 homo sapien
C 36	115	4.5	907	12 Q6E537	Q6E537 human herpe
C 37	115	4.5	961	3 Q92223	Q92223 emericella
C 38	115	4.5	1012	4 Q43393	Q43393 homo sapien
C 39	115	4.5	1012	4 Q73539	Q73539 homo sapien
C 40	115	4.5	1400	11 Q9ESU6	Q9ESU6 mus musculus
C 41	115	4.5	1400	11 Q9VHP8	Q9VHP8 mus musculus
C 42	114.5	4.5	1063	12 Q991E6	Q991E6 rubella vir
C 43	114.5	4.5	1063	12 Q991E4	Q991E4 rubella vir
C 44	114	4.5	585	12 Q41935	Q41935 murid herpe
C 45	114	4.5	810	11 Q60583	Q60583 mus musculus

ALIGNMENTS

RESULT 1
Q96109 PRELIMINARY; PRT; 207 AA.
AC Q96109
DT 01-DEC-2001 (TREMBLrel. 19, Created)
DT 01-DEC-2001 (TREMBLrel. 19, last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, last annotation update)
DE Unknown (Protein for IMAGE:4139786) (Fragment).
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCB1_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=MUSCLE;
RA Strausberg R.;
RL Submitted (MAY-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC007901; AAH07901.1; -
FT NON TER 1
SQ SEQUENCE 207 AA; 23189 MW; 74335A78000B16B0 CRC64;
Alignment Scores:
Pred. No.: 3.18e-39 Length: 207
Score: 519.50 Matches: 105
Percent Similarity: 59.66% Conservative: 0

Best Local Similarity: 59.66% Mismatches: 1
 Query Match: 20.56% Indels: 70
 DB: 4 Gaps: 1

US-09-989-919-15 (1-1397) x 096109 (1-207)

QY 2 GTGCTGACCTGTACCGGAGGGGAGTATCTGCAGAACTCCAGGCAAGCAGCAGTACC 61
 DB 102 ValLeuHisLeuYrHrGserGlyGlnYrLeuGlnHnSerThrAlaSerSerThr 121
 QY 62 GAGTACCACTGTATCCCAAGACAGACCATCCCGGAGAGACTACCGCTGTGGCCATCC 121
 DB 122 GluTyrGlnCysIleProHnSerThrIleProGlnGlnAspYrHrGysTrpProSer 141
 QY 122 TACCAACCAAGGAGGCTGCTCTTTCAGGTGTAACTGGCTGAGGCTGTGATGTCTGT 181
 DB 142 TyrHisHisGlySerCysLeuLeuSerValPheAsnLeuAlaGlnAlaValAspValCys 161
 QY 182 GAGAGCCATGCCAGGTGGGCTTGTGTGTCACCAACAGACCACTGCAGAGTGA 241
 DB 162 GluSerHisAlaGlnCysAlaGlnAlaPheValAlaThrAsnGlnThrTrpHrGlyArg 181
 QY 242 CCAGTGGAGAGACCCCTTCCAGAGAGATGGCAGACCTCTGTGAGTTGATAGTAGT 301
 DB 182 -Gln----- 182
 QY 302 GATCCCCCATCGAAGTCAGAGGGGGTGTGAGGTATGAGAGAGGTATACGTCTT 361
 DB 182 ----- 182
 QY 362 CAAGCAGTCAAATTAGGAGAAATGCTTGTCTCCAGAAAGAAACATCCAGCCCTGT 421
 DB 182 ----- 182
 QY 422 TACCTTCACCTCTGCCCCCAGAGTCGAGCTGTCTTTTCAAGCTGATGAGACCA 481
 DB 183 -----LeuValPhePheIleThrGlyTrpSerGly 192
 QY 482 AGTGTCTCTGATCCCAAGACCAATATGTGAAGGCTCTTGGC 527
 DB 192 nValValProAspProAsnLysThrThrValValLysAlaSerGly 207

RESULT 2
 ID 08VBZ6 PRELIMINARY; PRT; 268 AA;
 AC 08VBZ6;
 DT 01-MAR-2002 (TREMblrel) 20, Created)
 DT 01-MAR-2002 (TREMblrel) 20, Last sequence update)
 DT 01-JUN-2002 (TREMblrel) 21, Last annotation update)
 DE Hypoetical 30.1 kDa protein (Fragment).
 GN A115348.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=LIVER;
 RA Strassberg R.;
 RL Submitted (JAN-2002) to the EMBL/GenBank/DBJ databases.
 DR EMBL: BC022157; AAR22157.1;
 DR MGD: MGI:2146795; A115348.
 KW Hypoetical protein.
 FT NON TER 1
 SQ SEQUENCE 268 AA; 30090 MW; 688B6E2C929E35C CRC64;

Alignment Scores:
 Pred. No.: 4.81e-31 Length: 268
 Score: 431.00 Matches: 90
 Percent Similarity: 53.71% Conservative: 4
 Best Local Similarity: 51.43% Mismatches: 10
 Query Match: 17.06% Indels: 71
 DB: Gaps: 2

US-09-989-919-15 (1-1397) x 08VBZ6 (1-268)

QY 5 CTGACCTGTACCGGAGGGGAGTATCTGCAGAACTCCAGGCAAGCAGCAGTACCNG 64
 DB 165 LeuHisLeuPheArgSerGlyGlnYrLeuGlnHnSerThr---SerSerGlnAlaGln 183
 QY 65 TACCACTGTATCCCAAGACAGACCATCCCGGAGAGACTACCGCTGTGGCCATCC 124
 DB 184 TyrGlnAlaGlnIleProHnSerThrIleProGlnGlnAspYrHrGysTrpProSer 203
 QY 125 CACCAAGGAGGCTGCTCTTTCAGGTGTAACTGGCTGAGGCTGTGATGTCTGTGAG 184
 DB 204 HisHisGlyGlyCysLeuLeuSerValPheAsnLeuAlaGlnAlaIleAspValCysGln 223
 QY 185 AGCCATGCCCATGTGTGGGCTTGTGTGTCACCAACCAACCACTGCAGAGTGA 244
 DB 224 SerHisAlaGlnCysArgAlaPheValAlaThrAsnGlnThrTrpHrGlyArg-Ly 243
 QY 245 GTGGAGAAAGCCCTTCCAAAGAGATGGCAGACCTCTGTGAGTTGATAGTAGTANT 304
 DB 243 S----- 243
 QY 305 CCCCATCGAAGTCAGAGGGGGTGTGAGGTATGAGAGAGGTATACGTCTTCAA 364
 DB 243 ----- 243
 QY 365 GGCAGTCAAATTAGGAGAAATGCTTGTCTCCAGAAAGAAACATCCAGCCCTGTAC 424
 DB 243 ----- 243
 QY 425 CTGTACCTCTGCCCCCAGAGTCGAGCTGTCTTTTCAAGCTGATGAGACCACT 484
 DB 244 -----LeuValPhePheIleThrGlyTrpSerGlnVal 254
 QY 485 GGTCTCTGATCCCAAGACCAATATGTGAAGGCTCTTGGC 527
 DB 254 lValProAspAlaGlyLysThrThrValLysAlaProGly 268

RESULT 3
 ID 000512 PRELIMINARY; PRT; 1394 AA.
 AC 000512;
 DT 01-JUL-1997 (TREMblrel) 04, Created)
 DT 01-JUL-1997 (TREMblrel) 04, Last sequence update)
 DT 01-DEC-2001 (TREMblrel) 19, Last annotation update)
 DE B-cell CLL/Lymphoma 9.
 GN BCL9.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
 NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Willis T.G., ZALCBERG I.R., Coignet L.J.A., Wlodarska I., STUL M.,
 RA Jadayel D.M., Bastard C., Treleaven J.G., Catovsky D., Silva M.L.M.,
 RA Dyer M.J.S.;
 RT "Molecular cloning of translocation t(1;14) (q21;q32) defines a novel
 gene (BCL9) at chromosome 1q21.";
 RL Blood 91:1871-1873 (1998).
 DR EMBL: Y13620; CAA73942.1;
 KW B-cell.

Alignment Scores:
 Pred. No.: 0.000229 Length: 1394
 Score: 143.00 Matches: 118
 Percent Similarity: 31.03% Conservative: 39
 Best Local Similarity: 23.32% Mismatches: 161
 Query Match: 5.59% Indels: 188
 DB: Gaps: 24

US-09-989-919-15 (1-1397) x 000512 (1-1394)


```

QY 1221 GGGTTTCATGAAAGAAATGAGCAGAAAGGCCGCCGCCACCC----- 1174
DB 770 GYGLHHSIPROGLNGLN-----GluTYrGLYMeTGLYProArgProPheLeuPromet 787
QY 1173 GGGCCAGGCTCTGGCTCA----- 1156
DB 788 SerGInGLYProGLYSerAsnSerGLYLeuArgAsnLeuArgGLuProILeGLYProAsp 807
QY 1155 -----AACCAACCTGCTCCCTGACCCAGCTCTG-----GGTCAT 1120
DB 808 GlnArgThrAsnSerArgLeuSerHisMetProProLeuAsnProSerSerAsn 827
QY 1119 GGTAGAGGTGACCTCTGCCCCCTGACCCCTGGAGTGTCCCTGTGCGG 1060
DB 828 ProThrSerLeuAsnThrAlaProProValGlnArgGLYLeuGLYArgLysProLeuAsp 847
QY 1059 AGGTCAAGATGACCTCTCCAGGTACAGATCAACCATCTGAGACATCCCCGCCAA 1000
DB 848 IleSerValAlaGLYSer-GlnValHisSer-----ProGLY11LeAsnProLeuLys 864
QY 999 GTCTCTGTCCAGACCTAGCTAGAGAAACTCCAC----- 963
DB 864 SerPro-----ThrMetHisGlnValGlnSerProMetLeuGL 877
QY 962 -----GATGATTTGTCACAGCAGAGCCGCCAGCTCTGTGCGATGCA 916
DB 877 YSerProSerGLYAsnLeuLysSerProGlnThrProSerGlnLeu-AlaGLYMetLeuA 897
QY 915 GCAGCAGCAGCAGGGGCTTAGCTACATTTCTCACTATTGAAACAGATGATGGGCTGC 856
DB 897 laGLYProAlaAlaAlaAlaSerIle----- 905
QY 855 CCACGGAACCGTCCAGAGGAGGGGACAGCCCGCAGTGGGAGACTAGCTAGCTTGC 796
DB 906 -----LysSerProProVal-----LeuGLYSerAlaAla 916
QY 795 TACAGCCACAGGGTTGAGAGAGCAGTATACCCCATGATTC-----TGACTAGGA 742
DB 916 laSerProValHisLeuLysSerProSerLeuProAlaProSerProGLYTrpHisSers 936
QY 741 GCCAGAGTCCCATAGCTAGGCTGTGAACCTCACTTTTAACATTGGCATTTATTCACG 662
DB 936 exProLysProPro----- 940
QY 681 TTGTCTCTGTCATCTGTCTGTCTGGAGTTAGCTTTGTCTGACGGGTTCCAGGT 622
DB 941 -----LeuGlnSerProGLY11LeProProA 949
QY 621 GACATGACAGTCTCCAGTCAAGTCACTCCCTCC-----ACAGGCAAGCCAGCTGCGAG 565
DB 949 snHisLysAlaProLeuThrMetAlaSerProAlaMetLeuGLYAsnValGLY----- 966
QY 564 GATAGTCAAGTGTGACCGGAGCCCTCAGATAGTAGGACGAGGCGCTTCACATATGTG 505
DB 967 -----SerGLYGLYProProProProThrAlaSerGlnProAlaSerValAsnIleProG 985
QY 504 TCTTTGTGGATGAGGACCACTTGCTCCATCACTTTGAAAAGACCAAGCTGCCGAC 445
DB 985 LysLeuProSerSerThrProLysThrMetProProGLYProThrLysSerGlnAsn 1005
QY 444 CTGGGGGGGAGAGTAGAGGTAAAGGCGTGAATTTCTCTTTCTGA-----GGCAAGA 388
DB 1005 ro-----LeuSerIleMetMetSerArgMetSerLys 1016
QY 387 CCATTCCTCCCTAATTGACTGCTTGAAGACAGTATACCTCTCTCATCACTCA----- 331
DB 1016 heAlaMetProSerSerThr-ProLeuTYrHisAspAlaIleLysThrValAlaSerSer 1035
QY 330 -----GCACCCCTCTGACTTCGATGG-----GGATCA 301
DB 1036 AspAspAspSerProProAlaArgSerProAsnLeuProSerMetAsnAsnMetProGLY 1055

```

```

QY 300 CTATATATCAACTCCAGAGAGGTCTGCACTTC----- 265
DB 1056 MetGLY11LeAsnThrGLN-----AsnProAlaIleSerGLYProAsnProValAlaPromet 1074
QY 264 -----CTTGAAGAGGCTTCTCCACAGTGGCTCACTCTCCAGTGG----- 225
DB 1075 ProThrLeuSerProMetGLYMetThrGln-ProLeuSerHisSerAsnGlnMetProse 1094
QY 224 -----TCTGTTGTG 214
DB 1094 rProAsnAlaValGLYProAsnIleProProHisGLYValPrometGLYProGLYLeu 1114
QY 213 ACCCAAAAGCCCGACACTGGGCGATGCTCTCAAGACATCAACAGCTTACAGCAGTTG 154
DB 1114 tSerHisAsnProIleMetGLYHisGLYSerGlnGlnProPrometValProGLNGLY- 1133
QY 153 AACACTGAAAGAGGAGAGCTCCCGTGT----- 126
DB 1134 -----ArgMetGLYPhenProGlnGLYPhenProProValGlnSerProProGlnGln 1151
QY 125 -----GGTAGATGCGCAGCAGCGGTAGCTTCTGGGGGATGTCGTC 82
DB 1151 lProPheProHisAsnGLYProSerGLYGLYGLYSerPheProGLYGLY----- 1168
QY 81 TCTGGATACACTGTACT-----CGTACTCTGCTTGCCTGGAGTTCG 34
DB 1169 -MetGLYPhenProGLYGLYGLYProLeuGLYArgProSerAsnLeuProGlnSerSer 1188
QY 33 AGATCTGCTCC 24
DB 1188 aaSpAlaAla 1191

```

```

RESULT 4
ID Q9N1P1 PRELIMINARY, PRT, 631 AA.
AC Q9N1P1;
DT 01-OCT-2000 (TrEMBLrel, 15, Created)
DT 01-OCT-2000 (TrEMBLrel, 15, Last sequence update)
DE 01-JUN-2002 (TrEMBLrel, 21, Last annotation update)
DE Submaxillary mucin (Fragment).
GN BSM1.
OS Bos taurus (Bovine).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
OC Bovidae; Bovinae; Bos.
OX NCBI_TaxID=9913;
RN (1)
RP SEQUENCE FROM N.A.
RC STRAIN=BRED ANGUS;
RX MEDLINE=20223253; PubMed=10759843;
RA Jiang W., Gupta D., Gallagher D., Davis S., Bhavanandan V.P.;
RT "The central domain of bovine submaxillary mucin consists of over 50
RT tandem repeats of 329 amino acids: chromosomal localization of the
RT BSM1 gene and relations to ovine and porcine counterparts.";
RL Eur. J. Biochem. 267:2208-2217(2000).
DR EMBL: AF176427; AAF67278.1; -
DR InterPro: IPR000194; ATPase_a/bcentre.
DR PRINTS: PR01561; EDG8RECEPTOR.
DR PROSITE: PS00152; ATPASE_ALPHA_BETA; UNKNOWN_1.
FT NON TER 1
FT NON TER 631
SQ SEQUENCE 631 AA; 57067 MW; 184E9F0C2436115A CRC64;

```

Alignment Scores:

```

Pred. No.: 0.000982 Length: 631
Score: 135.50 Matches: 123
Percent Similarity: 31.88% Conservative: 30
Best local Similarity: 25.62% Mismatches: 147
Query Match: 5.29% Indels: 180
DB: 6 Gaps: 19

```

US-09-989-919-15 (1-1397) x Q9N1P1 (1-631)

```

QY 1191 GGGCCCCGCCCCCAGGCTCTGCTCAACACCACTGC----- 1144
DB 50 GlyGlyAlaValThrProAlaProSerProGlyGlySerSerAlaThrAlaGlyHisGly 69
QY 1143 -----TCCCTGACCCCACTTGGCTCAT-----GCTAGGCTGACCTCT 1102
DB 70 ValGlySerAlaThrValAlaGlnAlaSerGlyAlaThrGlyAlaAspValLeuArgSer 89
QY 1101 GCCCCCTGAC-----CCCTGGGCTGGAGTGTCCCTGCGGGAGGTGAGATAGACC 1045
DB 90 GlyThrSerLeuProValSerGlyValAlaValSerProGly-----Ser 104
QY 1044 TCCCGCAGTACAGAAATCACCATCTCGAGACATCCCGCCCAAGTCTCTGTCAGAC 985
DB 105 SerProGlyArgSerGlyAlaThrAla----- 113
QY 984 CTAGCTGAGAGAACCTCCAGCATGATGTTCCCAAGCAGCCCCCAGCTGTGTG 925
DB 114 -----ValSerSerGlnGlySerGlnProThrVal 123
QY 924 GCATGTGCAGACAGACA-----GCAGGGCTTACTACTACTCTCA 883
DB 124 AlaLeuSerGlyAlaThrGlyThrSerValGlyProSerGlyThrArg-PheSerSerSe 143
QY 882 CACTATTGAAACAGAGATGATGGGCTGCCAGGAGACCGTCCAGGGAGGGGACACA----- 828
DB 143 rAlaIleProAlaThrProGlySerThrThrGlyAlaGlyAlaGlyAlaGlyThrProG 163
QY 828 ----- 828
DB 163 yValaPserGlnGlnSerAlaSerLeuAlaAlaAlaArgProThrAlaLeuGlyPyr 183
QY 827 -----GCCCCAGTGGGGAGACTAGCTGAGCTTCTTACAGCCCAAGGTTGG 778
DB 183 oglyThrSerAlaProSerGlyGlnThrSer-GlnSerArgSerSerValProGly-- 201
QY 777 AGAGCAGTATACCCCATGATCTCTGACTAGAGAGCAGCAGCTCCATAGCTAGGCTG 718
DB 202 ---GlySerGlnThrGlnGlnProGlyThrGlySerGlnSerProThrLeuSerLeu 220
QY 717 ---GTAACCTACATTTTAACTTTGACATTGACATTGACGTTTCTGCTGACATCTGTC 661
DB 221 GlyValThrArgThrThrAlaLeuArgGlySerGlyThrArgValProSerThrGlyVal 240
QY 660 TGTCTGGAGTATAGCTTGTCTGACAGGGTTCACAGGTACATGACAGTCTCCAGTGC 601
DB 241 -----SerGlyLeuProGlySerThrGln-GlyGlySerAla 252
QY 600 AAGTCACT----- 593
DB 252 aAlaA-ThrGlyGlySerGlyAlaGlySerGlyProThrAlaProValSerGlyGlnThra 272
QY 592 -----CCCTCAGAGCAAGCCAGCTGCTGAGAT 562
DB 272 rGthrIleValIleSerGlyThrAsnValProValSerGlyAlaProValThrProGlys 292
QY 561 AGTCACTGCTGACCGG---AGCCCTCAGATAGTACAGCCAGAGGCTTACATATATGG 505
DB 292 exSerIaGlySerSerGlyAlaProGlyAlaGlyGlyProGlySerGlnThrAlaSerP 312
QY 504 TCTTGTGGGATAGGAGACACTGGCTCATCCAGCTTGAAGAAAGACAGCTGCCAG 445
DB 312 rIleuSerGlyAlaAlaGlyTh----- 319
QY 444 CTGGGGGGGAGAGGTAGAGATAGAGGCTGATGTTCTTCTTCTGAGGAGACACCA 385
DB 320 -----SerAlaThrGlySerGlyThri 327
QY 384 TTTCTCTAATTGACTGCTTGAAGACAGTATACTCTCTCATCACTAGACACCC 325
DB 327 IeIleProProSerGlyAla-----Pro-ValThrProGlnPro 319
QY 324 CCTCTACTTCGATGGGGGATCACTATATCACTCAGAGAGGTCTCTGCATCTGC 265

```

```

DB 340 ProIleIleSerThrGlyAlaSerAlaGlyProSerAlaSerSerGlnSerThrValThr 359
QY 264 CTGGAGAGGCTTTCTCCACTGCTTCACTGCTTCCAGTGCTGCTGTTGACACCAAG 205
DB 360 LeuProGlyAla-----ThrGlyThrAspVal----- 368
QY 204 GCCCCAGACTGGGATGCTCTCAGACATCCAGCTCAGCCAGGTTGAACACTGAA 145
DB 369 -----LeuArgSerGlyThrSerLeu-ProValSerGlyAlaValThr 383
QY 144 AGAGCAGCTCCCTGCTGTGATGAGATGGCCAGACCGGTACTTCTCGGGAGATG 85
DB 383 rProAlaSerSerProGlyGlySerSerAlaThrAlaGlyProAlaValGlySerGlnTh 403
QY 84 CTGTCTG-----GATACACTGTACTCTGCTGCTGCTGCGTGGAGTTCTGC 34
DB 403 rThrValGlnAlaSerGlyAlaThrGlyThr----- 413
QY 33 AGATACCTGCCGCTCGGT-----ACAGGTGACGA 3
DB 414 -AspValLeuArgSerGlyAlaSerLeuProValSerGlyAlaAla 428

RESULT 5
ID 075851 PRELIMINARY; PRT; 4123 AA.
AC 075851;
DT 01-NOV-1998 (TREMBLrel. 08, Created)
DT 01-NOV-1998 (TREMBLrel. 08, Last sequence update)
DE 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE WUGSC:H.DJ0751H13.1 protein (Fragment).
GN WUGSC:H.DJ0751H13.1
OS Homo sapiens (Human)
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Leonard S., Graves T., Stromatt C.;
RT "The sequence of Homo sapiens PAC clone RP4-751H13.";
RN Submitted (JUN-1998) to the EMBL/GenBank/DBD databases.
RA
RA SEQUENCE FROM N.A.
RA Waterston R.;
RL Submitted (DEC-1999) to the EMBL/GenBank/DBD databases.
DR EMBL; AC004877; AAC36301.1; --
DR HSSP; P01130; 1AUJ.
DR InterPro; IPR000923; BlueCu.1.
DR InterPro; IPR001064; Crystallin.
DR InterPro; IPR000561; EGF-like.
DR InterPro; IPR000421; FAS8.C.
DR InterPro; IPR001092; HLH_Basic.
DR InterPro; IPR001072; LIL.
DR InterPro; IPR001077; VWF_C.
DR InterPro; IPR001846; VWF_D.
DR Pfam; PF00754; F5_P8_type_C.1.
DR Pfam; PF00057; IId_recept_a.11.
DR Pfam; PF01826; TIL.5.
DR Pfam; PF00090; TSP.1.14.
DR Pfam; PF00094; VWD.3.
DR PRINTS; PR00261; LDRECEPTOR.
DR SMART; SM00231; FAS8C.1.
DR SMART; SM00192; LDla.10.
DR SMART; SM00209; TSP1.14.
DR SMART; SM00214; VWC.1.
DR SMART; SM00216; VWD.3.
DR PROSITE; PS00280; BPTI_KUNITZ_1; UNKNOWN_1.
DR PROSITE; PS00196; COPPER_BLUE; UNKNOWN_1.
DR PROSITE; PS00225; CRYSTALLIN_BETAGAMMA; UNKNOWN_1.
DR PROSITE; PS00022; EGF_1; UNKNOWN_1.

```


DR SMART; SM00317; SET; 1
 DR PROSITE; PS00398; RECOMBINASES_2; UNKNOWN_1
 DR PROSITE; PS00280; SET; 2
 SQ SEQUENCE 4957 AA; 531840 MW; 1026562E1419CE8D CRC64;

Alignment Scores:

Pred. No.: 0.00766 Length: 4957
 Score: 127.50 Matches: 106
 Percent Similarity: 33.97% Conservative: 36
 Best Local Similarity: 25.36% Mismatches: 149
 Query Match: 4.98% Indels: 128
 DB: 4 Gaps: 22

US-09-989-919-15 (1-1397) X 014687 (1-4957)

QY 1219 CTTTCATGGAAGCAATGNG-----GCAGAAAGCCCCGGGCC 1181
 DB 1765 LcnglyleuargProglIngluProProAlaaglAlaLeuAlaProserProProser 1784
 QY 1180 CACCCC-----CGCCCCAGTCCGTGCTCAACCAACAC-----CTGCTC 1142
 DB 1785 HisProaspIleheargProglYserTyThrasProTyAlaaglInProleuThr 1804
 QY 1141 CCTGACCCAGTCTTGCTCATGCTGAGTGCACCTCTGCCCT----- 1096
 DB 1805 ProArgProglIn-ProProProProglInserCySAlaLeuProProArgserleuPr 1824
 QY 1095 ----GACCCCTGGGCTGGCTGGAGTGTCTCCCTGTGGAGAGTCAGATGACCTCCCC 1040
 DB 1824 oSerAaPProPheSerArgValProValSerProglInserGlnSerSerInser 1844
 QY 1039 AGGTACAGATCAACCCACATCTTGAGACATCCCGCAAGTCTCCGTGCAGACCTAAG 980
 DB 1844 OleuthrProArgProleuSer-----AlaGlaAlaPheCyProser----- 1858
 QY 979 CTGAGAGAACACTCCAGATGATGTCTCCACAGACCCCGCTGCTGTGCATG 920
 DB 1859 ----ProValThProArgPhe-----GlnSerProAspPro----- 1869
 QY 919 TGCAGACAGACGACGAGGGGCTTACCTACATCTCACTATTGTAACACAGATGGG 860
 DB 1870 ----TyrSe 1871
 QY 859 CTGCCACGAGAACCTGCCAGAGAGG-----CACAGCCCCAGG 818
 DB 1871 rArgProProserArgProglInserArgAspProPheAlaProleuHislyProArg 1891
 QY 817 GGGAGACTAGCTGAGCTTGCTTACAGCCACAGGGGTGAGAGGAGT-----CATAC 764
 DB 1891 gProglInProProglInValAlaAlaheyls-----AlaGlySerleuAlaHisTh 1907
 QY 763 CCCCAGT---ATTCCTGAGCTAGAGACGACAGCTCCCATAGCTAG--GCTGTGTAATCAG 708
 DB 1907 rSerleuGlyAlaGlyPheProAlaAlaLeuProAlaGlyProAlaGlyLeuLeu-- 1926
 QY 707 ATTTTACATTGGCATTATTGACCTTGCTGTGCTGATCATCTGTCTGTGGAGATGT 648
 DB 1927 ----HisAlaIyValProserGlyglInProProAlaPheValArgSerProglYh 1944
 QY 647 AGCTTTGTGACGGGGTCCAGGTGACATGACAGTGCAGTCAAGTCACT----- 593
 DB 1944 rGlyAlaPheValGlyThProser-----PrometArgPheThrhePr 1959
 QY 592 ----CCCTCACAGGAGAGCCAGCCAGCTG 570
 DB 1959 oGlnAlaValGlyglInProserleuAlysProProValProglInProglYleuProProPr 1979
 QY 569 CTGAGGATAGTCAAGTGT---CAGCCGAGCCCTCAGATAGTCAAGCCAGGCTTAC 513
 DB 1979 oHisGlyIleAlaSerHisPheGlyProglYProThrleuGlyLysProglInserThras 1999
 QY 512 ATATGTGCTTGTGGATCAAGGACCACTGGCTCCATCCAGTCTTGAAGAACGACG 453

DB 1999 nTyThrValAlaThrGlyAsn-----PheHisPro-----Se 2010
 QY 452 CTGCCGACCTGGGGGACAGAGGTGAGAGTAAACAGGCTGATGTTCTTCTTGAGAG 393
 DB 2010 GlySerProLeuGlyProserSerGlySerThrGlyGlySerTyGlyLeuSer-Pro 2030
 QY 392 CAAGACATTCCTCCCAATTGACCTGCTTGAAGACAGCTATACCTCTCATACCT 333
 DB 2030 euArgProProserValLeu----- 2036
 QY 332 CAGACACCCCTCTGACTTCGATGGGAGATCACTATATCAACCTCCAGAGAGTCTG 273
 DB 2037 ----ProProProAlaProAsp---GlySerleuProTy---LeuSerHisGlyAla 2053
 QY 272 CCATCTCCTTGGAAAGGCTTCTCCACT-----G 243
 DB 2053 ergInArgserGlyIleThrSerProValGlyLysArgGlyLysProglYthrGlymeG 2073
 QY 242 GCTACCTGTCAGAGTGTCTGTTGTTGTCACCAACAGCCGACACTGGGATGCTCT 183
 DB 2073 lYSerSerleuAlaThrAlaGlyLeuProglYThrGlnAspProglYmetSerGlyLeu 2093
 QY 182 CACAGACATCCACAGGCTCCAGGCTTGTAACACTGAAGAGGAGGAGCTC 133
 DB 2093 ergInThrGlyLeuGlyLysGlnArgGlnArgGlnArgGlnArgGlnLeu 2109
 RESULT 7
 ID 014686 PRELIMINARY; PRT; 5262 AA.
 AC 014686;
 DT 01-JAN-1998 (TREMBLrel. 05, Created)
 DT 01-JAN-1998 (TREMBLrel. 05, Last sequence update)
 DT 01-MAR-2002 (TREMBLrel. 20, Last annotation update)
 DE ALR.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=97388474; PubMed=9247308;
 RA Prasad R., Zhadanov A.B., Sedkov Y., Bullrich F., Druck T.,
 RA Rallapalli R., Yano T., Alder H., Croce C.M., Huebner K., Mazo A.,
 RA Canaan E.
 RT "Structure and expression pattern of human ALR, a novel gene with
 RT strong homology to ALR-1 involved in acute leukemia and to Drosophila
 RT trithorax."
 RT Oncogene 15:549-560 (1997).
 CC -1- SIMILARITY: CONTAINS 1 SET DOMAIN.
 DR EMBL; AF010403; AAC51734.1. -
 DR InterPro; IPR003889; FYRIC.1.
 DR InterPro; IPR003888; FYRIC.N.
 DR InterPro; IPR000910; HMG_12_box.
 DR InterPro; IPR003616; PostSET.
 DR InterPro; IPR002965; P-rich_extensn.
 DR InterPro; IPR001822; Recombinase.
 DR InterPro; IPR001214; SET.
 DR InterPro; IPR001965; ZnF_PHD.
 DR Pfam; PF00628; PHD; 5.
 DR Pfam; PF00856; SET; 1.
 DR PRINTS; PR01217; PRICHEXTENSN.
 DR SMART; SM00542; FYRC; 1.
 DR SMART; SM00541; FYRN; 1.
 DR SMART; SM00398; HMG; 1.
 DR SMART; SM00249; PHD; 7.
 DR SMART; SM00508; PostSET; 1.
 DR SMART; SM00184; RING; 3.
 DR SMART; SM00317; SET; 1.
 DR PROSITE; PS00398; RECOMBINASES_2; UNKNOWN_1.
 DR PROSITE; PS00280; SET; 2.
 SQ SEQUENCE 5262 AA; 564171 MW; 26B7C74CAD417E44 CRC64;


```

Db 193 -----ValSerGlyGlnGlySerGln----- 199
QY 951 CCCAGCAGCCGCCCGCTGCTGGCATGTGCAGCAGCAGCAGCGGCTTAGCT 892
Db 200 ProThrValAlaLeuSerGlyAlaThnGlyThrSerAlaGlyProSerGlyThrArgSer 219
QY 891 ACATTCTCACATTTTGAACACAGTAGTGGGCTGCCACGGAACCGTCCAGGAGGG 832
Db 220 Thr-SerSerAlaAlaProAlaThrProGlySerThrThnGlyArgAlaAlaGlyAlaG 239
QY 831 CACA----- 828
Db 239 YThrProAlaValAlaSpSerGlnThnAlaSerLeuProAlaAlaAlaArgProThrAl 259
QY 827 -----GCCCCAGTGGGAGACTAGCGTGGCTTACAGCCCA 787
Db 259 AlaGlyProGlyThrSerAlaProSerGlyGlnThnSer--GluSerArgSerSerVal 278
QY 786 CAGGTTGGAGAGCAGCTATACCCCATGATTCCTGGATAGAGCCAGCAGTCCCAT 727
Db 279 ProGly-----GlySerGlnThrThnGlnProGlyAlaGlySerGluSerProThr 296
QY 726 GCTA-----GGCTGGTAAACTCACAATTTTAA 700
Db 297 -LeuSerProGlyAlaThrArgThrThrAlaLeuArgGlySerGlnThrArgValProse 316
QY 699 ATTGGCATTATTGCAGCTTTGCTGTCATC----- 665
Db 316 rThnGly-----ValSerGlyLeu-ProGlySerThnGlnGlySerAlaAlaThnGlyG 335
QY 664 -----TGTCTGTGGAGATGTACTTTGTCTGGAGGGTCCAGGTGCAGTGCAG 613
Db 335 LysSerGlyAlaGlySerGlyProThrAlaProValSerGlyGlnThrArgThrSerVal 355
QY 612 TGCTGCAGTGCAGTCACTCCCTCCAGCAGCCAGCCAGCTGCTGAGTACAGCTG 553
Db 355 LeuSerGlyThrAlaVal-----ProValSerGlyAlaProValThnProGlySerSerAlaG 374
QY 552 GTCAGCGC---AGCCCTCAGATAGTGCAGCCAGAGCCCTTCACATATGCTGCTG--- 500
Db 374 LysSerSerGlyAlaProGlyThrGlyGlyProGlySerGlnThrAlaSerProLeuSerG 394
QY 499 -----TTGGATCAGGAGAC----- 485
Db 394 LysAlaAlaGlyThrSerAlaThrThrLysSerGlyThrSerLleProProSerGlyAlaProV 414
QY 484 --ACTGCTCCATCAGTCTTG-----A 463
Db 414 AlThrProGlnProProLeuLleSerThrGlyAlaSerAlaGlyProProAlaSerSerG 434
QY 462 AAAAGACCACTGCCAGCTGGGGGAGAGAGTAGAGTAACAGGGCTGATTTTCTC 403
Db 434 LysSerThrValThrLeuProGly-----AlaThnGlyThrAspValLeuA 449
QY 402 TTCTTGAGGAGAACACATTCCTCAATTGATGCTTGAACACGTAATACCTCTCT 343
Db 449 rGserGly-----ThrSerLeuProValS 457
QY 342 CTCATCACTCAGACCCCTCTGACTTCGATGGGGATCATTATCACTGCTCAG 283
Db 457 ergLysGlyAlaValThrProAla-ProSerProGlyLysSer-----SerAlaThr 473
QY 282 AGAGTCTCT-----GCCATCTCCCTTGAAGGGCTTTCACCACTGGCTCAGCT 235
Db 474 AlaGlyProGlyValGlySerAlaThrThrValGlnAlaSerGlyAlaThnGlyAlaSp 493
QY 234 GTCACAGGTGCTGCTGTTGTGATCACCAAAAGCCGACACTGGGATGCTCTCAGACACA 175
Db 494 Val-----LeuArgSerGlyThr 499
QY 174 TCACAGCCTCAGCAGGTTGAACACTGAAGAGGAGCTCCCGTGT----- 126

```

```

Db 500 SerLeu-ProValSerGlyValAlaValaSerProGlySerSerProGlyArgSerGlyAl 519
QY 125 -GTTAGATGGCCAGCAGCGGATCTTCC-----GGGGATGGNG 85
Db 519 aThrAlaValSerSerGlnGlySerGlnProThrValAlaLeuSerGlyAlaThnGlyTh 539
QY 84 CTGCTGGATATCAGTCTGATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 25
Db 539 rSerValGlyProSerGlyThnArgPheSer-----SerSerAlaAlaProAl 555
QY 24 CCGTCCCGT-----ACAGGTGACGA 3
Db 555 aThrProGlySerThrThnGlyArgAla 564

RESULT 9
Q9EPX2 PRELIMINARY; PRT; 1280 AA.
ID Q9EPX2;
AC Q9EPX2;
DT 01-MAR-2001 (TRMBLrel. 16, Created)
DT 01-MAR-2001 (TRMBLrel. 16, Last sequence update)
DT 01-JUN-2002 (TRMBLrel. 21, Last annotation update)
DE Papilin.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CD-1;
RA Kramarova I.A., Kusachi N., Nelson R.E., Fessler L.I., Chen Y.,
RA Kramarova A.A., Kusachi-Gullberg M., Kramer J.M., Ackley B.D.,
RA Stieren A.L., Prockop D.J., Fessler J.H.;
RT "Papilin in development: a pericellular protein with a homology to the
RT ADAMTS metalloproteinase."
RL Development 127:5475-5485 (2000).
DR EMBL; AF314171; MAG41980.1; -.
DR HSSP; P12111; 2KMT.
DR InterPro; IPR000183; Decarboxylase2.
DR InterPro; IPR003598; Ig.
DR InterPro; IPR003598; Ig_C2.
DR InterPro; IPR003600; Ig_1like.
DR InterPro; IPR003006; Ig_MHC.
DR InterPro; IPR002223; Kunitz_BPTI.
DR InterPro; IPR000884; TSP1.
DR Pfam; PF00047; Ig_3.
DR Pfam; PF00014; Kunitz_BPTI; 1.
DR Pfam; PF00090; tsp_1; 5.
DR PRINTS; PR00759; BASICPTASE.
DR ProDom; PD000222; Kunitz_BPTI; 1.
DR SMART; SM00409; Ig_3.
DR SMART; SM00408; IGC2; 3.
DR SMART; SM00410; Ig_1like; 2.
DR SMART; SM00131; Kof_1.
DR SMART; SM00209; TSP1; 5.
DR PROSITE; PS00280; BPTI_KUNITZ_1; 1.
DR PROSITE; PS0279; BPTI_KUNITZ_2; 1.
DR PROSITE; PS00879; ODR_DC_2.2; UNKNOWN_1.
DR PROSITE; PS50092; TSP1; 3.
KW Immunoglobulin domain; serine protease inhibitor.
SQ SSQUNCE 1280 AA; 138824 MW; AE287705B561AF30 CRC64;

Alignment Scores:
Pred. No.: 0.0115 Length: 1280
Score: 124.50 Matches: 128
Percent Similarity: 30.348 Conservative: 41
Best Local Similarity: 22.988 Mismatches: 183
Query Match: 4.93% Indels: 206
DB: 11 Gaps: 31

US-09-989-919-15 (1-1397) x Q9EPX2 (1-1280)
QY 3 TGCTGCAGCTGTACCGAGAGGAGAGTATCTGCAGACTCCAGCAGGACAGCATGACG 62

```

[illegible]

D	b		723	rglucyargsersearAtrPheglYcyscytYAsrAnvalAlaserAlalaglPr	733					
Q	y		818	CACtgggggGctgtgcCCTCCCTGGAGCGGTTCCTGGGCAGCCCCCACTGttTCA	877					
D	b		733	OleuglY-----GluglYcysValglYalnProserTYralatYr---	746					
Q	y		878	TAGTGATGAATTAgACTAAAGCCCCCTGTCTGTGTGCACATGCCAGCAGAGCGGT	937					
D	b		747	-----ProvalnrgCyaleuNeurProSerAlaglnGlYserCY	759					
Q	y		938	GCGG-----GCTGCGTG-----GGACAATCCATCCGTGAGAGTGTCTCACGTT	982					
D	b		759	egLYasPrtrAlalalagrtPrYrhnevalAlaserValolYargcs-----	775					
Q	y		983	AGTCTGGACAGAGACTTGGCGGGGAGTGTCCAGA-----	1020					
D	b		776	-----AsnrgrRhetrrYrglYglYcysHnIe-glyASnalasnPhelaserG	793					
Q	y		1021	-----TGrngtgATTCtGTnAcTCTGGGAGGCTATCTGTACCTCCAGCGGGA-	1071					
D	b		793	IuglngluCYsMeelasnThrCYs-----ArglYg	803					
Q	y		1072	--CACTCCCGACGCCAGCCCA-----GGGTCAGAGGCGAGGTCACACCTCGACATGA	1123					
D	b		803	InhlSeglYrArgrArgrProglualaglyAlalaglyhnlarAlahlsval-----	820					
Q	y		1124	GCCAAGACTGGGTCAGAGGAGCAGGTGTGTTTAGCCAGGACCCTGGGCGGGGTGG-	1182					
D	b		821	-----AsprglYglYlnhrg-----glyProglYglY	830					
Q	y		1183	-----GCGGGGCTTTCGCTCATTTGCTTCATTAAG	1219					
D	b		830	IuglngluProaspTrphIsarAlaglyAlathrlleProargLeuProserProserG	850					
Q	y		1220	CCTCA-----AAGAGCGCAAAACCAGGCTTCCCCT	1251					
D	b		850	lyserProtrparYrArgrgluglngluProalaprOglYgluPropro	865					
<hr/>										
RESULT 10										
ID	075441	PRELIMINARY;	PRT:	669 AA.						
AC	075441:									
DT	01-NOV-1998	(TREMBLrel. 08,	Created)							
DT	01-NOV-1998	(TREMBLrel. 08,	Last sequence update)							
DT	01-MAR-2002	(TREMBLrel. 20,	Last annotation update)							
DE	Latent transforming growth factor-beta binding protein 4									
DE	(Fragment).									
OC	Homo sapiens (Human).									
OC	Eumaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;									
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.									
OX	NCBI_Taxid=9606;									
RN	[1]									
RP	SEQUENCE FROM N.A.									
RX	MEDLINE=98325059;	PubMed=9660815;								
RA	Sahrinen J., Taipale J., Momi O., Keski-Oja J.;									
RT	"Identification and characterization of a new latent transforming									
RT	growth factor-binding protein, LTBP-4."									
RL	J. Biol. Chem. 273:18459-18469(1998).									
DR	EMBL; AF054502; AAC93882.1;									
DR	HSSP; P35555; IEMN.									
DR	InterPro; IPRO00152;	Aex_hydroxyl.								
DR	InterPro; IPRO00561;	EGF-like.								
DR	InterPro; IPRO01881;	EGF Ca.								
DR	InterPro; IPRO02212;	Fibril-assoc.								
DR	Pfam; PF00008;	EGF; 6.								
DR	Pfam; PF00683;	TB; 2.								
DR	SMART; SM00179;	EGF_CA; 6.								
DR	SMART; SM00001;	EGF_like; 2.								
DR	PROSITE; PS00010;	ASX_HYDROXYL; 6.								
DR	PROSITE; PS01186;	EGF_2; UNKNOWN_2.								
KM	PROSITE; PS01187;	EGF_CA; 8.								
FT	Calcium-binding; EGF-like domain; Glycoprotein; Hydroxylation; Repeat.									
FT	NON_TER	1								

FT NON TER 669 669
 SQ SEQUENCE 669 AA; 71702 MW; 2C76C4AB114284 CRC64;
 Alignment Scores:
 Pred. No.: 0.0127 Length: 669
 Score: 123.50 Matches: 109
 Percent Similarity: 30.97% Conservative: 31
 Best Local Similarity: 24.12% Mismatches: 129
 Query Match: 4.89% Indels: 184
 DB: 4 Gaps: 27
 US-09-989-919-15 (1-1397) x 075441 (1-669)
 QY 110 TGCTGGCATTCACCAACGAGG-----AGCTGCTCTTTCAGTTTCAC 157
 DB 128 CysAlaProGlyIyrhArgProGlyProAlaGlyAlaSerCysLeu----- 142
 QY 158 CTGGCTGAGGCTGTGTCTGTGAGAGCCATGCCAGTGGGCTTTGTGTCACC 217
 DB 143 -----AspValAspGluCysSerGluGluAspLeuGlySerGlnSerGlyIleCysThr 159
 QY 218 AACACAGCACCTGAGCAGGTGAGCCAGTGGAGAACCCCTTCCAGGAGATGGCAGA 277
 DB 160 Asn---ThraspGlySerPheGluCysIle-----CysProProGlyIyrhArgAlaGly 176
 QY 278 CCT---CTGTGAGGTTGATAGATGATGATCCCTCCATCGAAGTCAAGGGGGTGTGAG 334
 DB 177 ProAspLeuAlaSerCysLeuAspValAspGlu----- 187
 QY 335 GTGATGAGAGAGAGGTATACGTGCTTCAAGCAGTCAAAATTAGAGAGATGGTTCGCC 394
 DB 188 ---CysArgGluArgGlyProAlaLeuGlySerGlnArgCysGluAsnSerProGly 206
 QY 395 TCAGAAAG-----AGAAACATCCAGCC----- 418
 DB 207 SerTyrArgCysValAlaArgAspCysAspProGlyTyrHisAlaGlyProGluGlyThrCys 226
 QY 418 ----- 418
 DB 227 AspAspValAsnGluCysGluThrLeuGlnGlyValGlySerGlyAlaAlaLeuCysGluAsn 246
 QY 419 -----TGTACTCTCACT----- 433
 DB 247 ValGluGlySerPheLeuGlyValCysProAsnSerProGluGluThrAspPromerThr 266
 QY 434 -----CTGCCCCCAGGCTGGCAGCTGGTCTTTTCAAGACTGATGAGCCAACT 484
 DB 267 GlyArgCysValProProArgThrSerAlaGlyMetPheProGly---SerGlnProGln 285
 QY 485 GGTCCCTGATCCCAACAAGACACATATGTGAAGGCTCTGGCTGACTTATCTGAGGGCT 544
 DB 286 AlaProAlaSer-----ProValLeuProAla 294
 QY 545 CGG---CTGACCAAGCTACTATCTCAGCAGCTGGCTGGCTGTGAGGAGAGACTTG 601
 DB 295 ArgProPro-PropProLeuProAlaGlyArgProSerThrProArgGlnGlyProValGln 314
 QY 602 CACTGGCAGACCTGATGATCACTGGAAGCCCTGACAGACAAAGCTAAATCCAGACAG 661
 DB 314 ySerGlyArgArgGluCysTyrPheAspThrAlaAla-----ProAsp-- 328
 QY 662 ACAGATGTGACAGGACAAAGCTGCATATGTCAAATGTTAAATGTGAGTTTACACAG 721
 DB 329 -----AlaCysAspAsnIleLeuAlaArgAsnValThr----- 339
 QY 722 CTAGCTATGG---GACTGCTGGCTCCCTAGTCCAGGAATCAATGAGGGGTATGACTGCTCTC 778
 DB 340 -----TTrpGlnGluCysCysThrValGlyGlnGlyTyrPoli----- 352
 QY 779 CAACCTGTGGGCTGTAGAGCAAGCTAGCTTCCCACTGGGGGCTGTGCCCTGCC 838
 DB 353 -----SerGlyCysArgIleGlnGlnCys-----Pr 361

QY 839 CTGGACGCTTCGTRGGAGCAGCCCATCACTGTTCATTAAGTGAATAGACTTAA 898
 DB 361 o-GlyThrGluThrAlaGluTyrGlnSerLeuCysPro----- 373
 QY 899 GCCCTGTGCTGCTGCTGCTGACATGCCACAGAGCGGTGG----- 939
 DB 374 -----HisGlyArgGlyTyrLeuAlaProSerGlyAspL 385
 QY 940 ---CGCTGGTGGGACATTCATGCTGAGAGTGTCTCTCACTTAAGTCTGG---ACAG 994
 DB 385 eusSerLeuArgArgAspValAspGluCysGlnLeuPheArgAspGlnValCysLysSerG 405
 QY 995 GAGACTTGGCGGGGAGTGTCCAGATGTGGGTGATCTGTACTGAGGAGGCTATCTCT 1054
 DB 405 lValCysValAsnThrAlaProGlyTyrSerCysTyrCysSerAsnGlyTyrTyr- 424
 QY 1055 GACCTCCGACAGGAGGACCTCCAGGC----- 1082
 DB 425 -----HisThrGlnArgLeuGluCysIleAspAsnAspGluCysAla 439
 QY 1083 ---CAGCCAGGAGGTCAAGGAGGAGGAGTGCACA-----CCTCAGCAT----- 1121
 DB 439 spGluGluProAlaCysGluGlyGlyArgCysValAsnThrValGlySerTyrHisCysT 459
 QY 1122 ---GAGCCAGACTG-----GGGTCAAGGAGCAGGTGT----- 1151
 DB 459 hrcCysGluProProLeuValLeuAspGlySerGlnArgArgCysValSerAsnGluSerG 479
 QY 1152 ---GGTTGAGCCAGACTGTGGGCGGGGGTGG 1181
 DB 479 InsSerLeuAspAspAsnLeuGlyValCysTyr 489
 RESULT 11
 ID Q9ARH1 PRELIMINARY; PRT; 647 AA.
 AC Q9ARH1;
 DT 01-JUN-2001 (TrEMBLrel. 17, Created)
 DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
 DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
 DE Receptor protein kinase PERK1.
 OS Brassica napus (Rape).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
 OC eurosids II; Brassicales; Brassicaceae; Brassica.
 OX NCBI_TaxID=3708;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Silva N.F., Goring D.R.;
 RT "Characterization of a novel Brassica napus receptor protein kinase
 RT PERK1."
 RC Submitted (MAR-2001) to the EMBL/GenBank/DBJ databases.
 CC -1- SIMILARITY: BELONGS TO THE SER/THR FAMILY OF PROTEIN KINASES.
 DR EMBL; AY028699; AAK21965.1; -
 DR InterPro; IPR000719; Euk_pkinase.
 DR InterPro; IPR002965; P-rich_extensn.
 DR InterPro; IPR002290; Ser_thr_pkinase.
 DR InterPro; IPR001245; Tyr_pkinase.
 DR Pfam; PF00069; pkinase; 1.
 DR PRINTS; PR01217; PRICHEXTENS.
 DR PRINTS; PR00109; TYRKINASE.
 DR Prodom; PD000001; Euk_pkinase.
 DR SMART; SM00220; S_TKc; 1.
 DR SMART; SM00219; TYRKc; 1.
 DR PROSITE; PS00107; PROTEIN KINASE_ATP; 1.
 DR PROSITE; PS00108; PROTEIN KINASE_DOM; 1.
 DR PROSITE; PS00108; PROTEIN KINASE_ST; 1.
 KW ATP-binding; Kinase; Serine/threonine-protein kinase; Transferase.
 SQ SEQUENCE 647 AA; 68482 MW; A69A8A0C0D5CA203 CRC64;
 Alignment Scores:
 Pred. No.: 0.0156 Length: 647
 Score: 122.50 Matches: 91
 Percent Similarity: 31.28% Conservative: 36

US-09-989-919-15 (1-1397) x Q9UPX0 (1-1400)

QY 1191 GGGCCCGGCCCCACCC-----GCCCGAGTCTGGCTCAACACACACTGCTCCTGACC 1135
 DB 954 GlyserylProgluProleuAlaProSerPro-----ValSer 966

QY 1134 CCACTTGGCTATGCTGAGGTGTGACCTTCCCTGACCCCTGAGGCTGGCTGGGA 1075
 DB 967 ProThrPhePro-----ProSerSerProSerAspTrpPro-----GlnGlu 980

QY 1074 GTGTCCCTGCTGGAGATGACAGATAGCTCCCGAGTACAGATACACCCACATCC--- 1018
 DB 981 ArgSerProGlyGlnHisSerAspGlyAlaSerProArgSerProValProThrThrLeu 1000

QY 1017 -----TGAGCATCCCGCCCAAGTCTCTGTCCAGACTTACG 979
 DB 1001 ProGlyLeuArgHisAlaProTrpGlnGlyProArgGly-----ProProAspSer 1017

QY 978 TGAGAGAACACTCCACATGATGTGCCACGAG-----CCC 940
 DB 1018 ProAspGlySerProLeuThrProValProSerGlnMetProTrpLeuValAlaSerPro 1037

QY 939 CCAACGGCTGCTGGCATGTGACAGACAGACAGAGGCTTACCTTACATCTCTCAC 880
 DB 1038 GluProPro-----GlnSerSerProThrProAlaPheProLeuAlaAlaSer 1053

QY 879 TATTGAACACAGTATGGGCTGCCA-----CGAACCTCCCGAGGAG 835
 DB 1054 TyrAspThrAsnGlyLeuSerGlnProProLeuProGlyAlaSerGlnLeuProGlyPro 1073

QY 834 GGGCACAGCCCC---CAGTGGGAGACTAGCTTACCTTACAGCCCAAGGGTTGG 778
 DB 1074 GlyGlnGlnProGlyProTrpGly-----ProGlnGlnAlaSerSerProAla----- 1089

QY 777 AGAGGAGTATACCCCATGATCTCTGACTAGACAGACAGACTCCCATATAGTAGCTG 718
 DB 1090 ArgGly----- 1091

QY 717 GTAACTCACATTTTAACTTGGCATTTATGACGTTTCTCTGTCACATCTGTCTGT 658
 DB 1092 -----IleSerHisIleValThrPheAlaProLeuLeuSer----- 1103

QY 657 CTGGAGTGTAGCTTGTCTGACAGGGTCCCGAGGTACATGACAGTGTCTGCCAGTGA 598
 DB 1104 -----Asp 1104

QY 597 TCACTCCCTCCACAGGACCCAGCTGCTGAGATAGTACGTGTACGCCGACCTC 538
 DB 1105 AsnValProGlnMetProGluProProThrGlnGlnSerGlnSerAsnValLysPheVal 1124

QY 537 AGATAGGTACAGCCAGAGGCTTACATATGTGTTGTGGATGACAGGACCACTTGGC 478
 DB 1125 GlnAspThrSerLys-----PheTrp-TyrLysProHisLeuSe 1137

QY 477 TCCATCC-----AGCTTGAAGAAAGACACTGCCACCTGGGGGGGACAGAGTGAG 427
 DB 1137 TArgAspGlnAlaIleAlaLeuLeuLysAspLysAspPro----- 1150

QY 426 AGTAAACAGGGCTGAGTGTCTCTTCTGAGAGCAAGACCTTCCCTAATTGACTG 367
 DB 1151 -----GlyAlaPheLeuIle--ArgAspSerHisSer----- 1160

QY 366 CCTTGAAGACACGATACCTCTCTCTATATCACTCAGACACCCCTGTGACTTCGATGG 307
 DB 1161 PheGlnGlyAlaLys--GlyLeuAlaLeuLysValAlaThrPro----- 1174

QY 306 GGATCACTATATCAACCTCCAGAGAGGTCTGACATCTCCCTTGAAGAGGC----- 254
 DB 1175 -----ProProSerAlaGln-----ProTrpLysGlyAspProVal 1186

QY 253 -----TTCCTCCACTGGCTC-----ACCTGTCCAGAGGT-----CTGGTTG 217

DB 1186 aIGluGlnLeuValArgHisPheLeuIleGluThrGlyProLysGlyValLysIleLysG 1206

QY 216 GTGACCAACAAGCCCGACACTGGGC-----ATGGCTCTCACAACATCCACAGCTCA 163
 DB 1206 LysProSerGlnProLysPheGlySerLeuSerAlaLeuValSerGlnHisSerLys 1226

QY 162 GCCAGGT-----GAACACTGAA 145
 DB 1226 eIFProLleSerLeuProCysCysLeuArgIleLeuSerLysAspProLeuGluGluThrP 1246

QY 144 AGAGGACACTTCCGTGTGTGAGATGGCCAGACAGCGGTACTTCTTGGGGAGTGG 85
 DB 1246 roGlu--AlaProValProThrAsnMetSerThrAlaAlaAspLeuValArgGlnGly 1265

QY 84 CTGTCTGGGATACACTGGTACTC 62
 DB 1265 laAlaCysSerValLeuThrLeu 1272

RESULT 13
 Q9UPX0
 ID Q9UPX0 PRELIMINARY; PRT; 763 AA.
 AC Q9UPX0;
 DT 01-MAY-2000 (TRENBLREL 13, Created)
 DT 01-MAY-2000 (TRENBLREL 13, Last sequence update)
 DT 01-DEC-2001 (TRENBLREL 19, Last annotation update)
 DE KIAA1030: protein (Fragment).
 GN KIAA1030.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=BRAIN;
 RX MEDLINE=99397452; PubMed=10470851;
 RA Kikuno R., Nagase T., Ishikawa K., Hirosewa M., Miyajima N.,
 RA Tanaka A., Korani H., Nomura N., Ohara O.;
 RT "Prediction of the coding sequences of unidentified human genes. XIV.
 RT The complete sequences of 100 new cDNA clones from brain which code
 RT for large proteins in vitro.";
 RL DNA Res. 6:197-205 (1999).
 DR EMBL; AB028953; BAA82982.1;
 DR InterPro; IPR003961; FN_III.
 DR Pfam; PF00041; fn3; 1.
 DR SMART; SM00060; FN3; 1.
 FT NON TER 1
 SQ SEQUENCE 763 AA; 82322 MW; B9756B33F2F2B619 CRC64;

Alignment Scores:
 pred. No.: 0.0245 Length: 763
 Score: 120.50 Matches: 84
 Percent Similarity: 31.40% Conservative: 35
 Best Local Similarity: 22.16% Mismatches: 134
 Query Match: 4.71% Indels: 126
 DB: 4 Gaps: 16

US-09-989-919-15 (1-1397) x Q9UPX0 (1-763)

QY 1182 CCCACCCCGCCCCAGTCTCT-----GGCTCAAC 1153
 DB 414 ProLeuProThrGlnGlyProPheGlnHisProThrIleProGluGluLysGlnAsn 433

QY 1152 CACACTGCTCTCTGACCCCAAGCTTGGCTCATGCTGAGGTGACACTTGGCCCTGAC 1093
 DB 434 AlaSerAsnSerThrLeuPro--LeuThrGlnThrProThrGlyGlyArgSerProGlu 452

QY 1092 CCCTGGGCTGGCTGGGAGTGTCCCTGCGGAGGTGACAGATACCTCCACAGGTACA 1033
 DB 453 ProTrp-----GlyArgProGluPheProPheGlnGlyLeu 464

QY 1032 GAATCACCAACATCTGAGACATCCCGCAAGTCTCTGTCCAGACCTAAGCTGAGAG 973
 DB 465 GluThrPro--AlaMetMetPheProHisGlnLeuProProCysAspValProGluSer 483

Db 91 AspThrIleIleAlaSerAsnProPheAspSerProGlnValSerAlaMetSerSer 110
 QY 1088 ----- 1088
 Db 111 SerAlaAlaAlaMetAlaAlaMetAsnGlnMetGlyGlyProGlyGlyGlnHis 130
 QY 1087 -----GGTGGCCTGGAGAGTCCCTG-----TCGGGA 1059
 Db 131 PheGlyGlyGlyProGlyGlyGlnHisProHisIleProLysPargMetGlyMetGlyGly 150
 QY 1058 GGTCAAGATAGTCTCCCAAGTACAGAAATCACCCATCTTGAGATATCCCGCCCAAG 999
 Db 151 GlyProProProProHisMetHisProHisMetHisProHisHisPro----- 167
 QY 998 TCTCTGTCTCAGACTTAAGTGAAGAACATCCAGATGATTTGTCACCGCAGCCCG 939
 Db 168 -----GlyGlyProMetGlyHisProHisGlyProHis 178
 QY 938 CACCGCTGTGTGGCATGTGCAGACAGACAGAGGAGGCTTTAGTCACTTCTCACT 879
 Db 178 sProHisMetGlyGly-----ProProProMetArgGlyMetSer----- 191
 QY 878 ATTGAACACATGATGGGGCTGCCACAGAACCGTCCAGAGAGGGACACAGCCCGAGT 819
 Db 192 -----ProMetHisProHisGlnMetGlyProGlyProGlyVal 204
 QY 818 GGGGAGACTAGCTGAGCTTGTACAGCCACAGGCTTGAGAGGAGT----- 769
 Db 204 IGIyleuPro-ProHisMetAsnHisGlyArgProGlyGlyProGlyGlyProGlyGlyP 224
 QY 768 -----CATACCCCATG-----ATTCTGACTAGAGCCAGCAGTCCCATAG 726
 Db 224 roValProMetGlySerProMetGlyGlyIleAlaGlyMetGlyGlyMetSerProMetG 244
 QY 725 CTNAGCTGGTAACTACATTTTAACATTGGCATTTTGCAGCTTTGCTGTGTCAT 666
 Db 244 Iy-Gly-----MetGlyGlyProSer-IleSerProHisHisMe 256
 QY 665 CTGTCTGTGGGATGTAGCTTTGTCTGAGGGGTT-----CCGAGTGACA 618
 Db 256 tGIyleMetGlyGlyLeuSerProMetGlyGlyGlyProAsnGlyProAsnProArg----- 274
 QY 617 TGCAGTGTCCAGTGCAGTCACTCCCTCCACAGAGCCAGCTGTGAGGATAGTC 558
 Db 275 -----AlaMetGlnGlySerProMetGlyGly----- 283
 QY 557 AGCTGTACGCCAGCTTCAGATAGTCAAGCAGAGGCTTCACATATGTGCTTTGTT 498
 Db 284 -ProGlyGlnAsnSerProMetAsnSerLeuProMetGly-SerProMetGlyAsn----- 301
 QY 497 GGGATCAGGAGCACTGGTCCATCCAGTCTTGAAAAAGACCAAGCTGCCGACTGGGGG 438
 Db 302 -----ProIleGlySer-----ProLeuGlyProProSerG 312
 QY 437 GCAGAGGTGAGAGGTACAGGGCTGATGTTTCTTTCTGAGAGCAAGACATTCCTCC 378
 Db 312 IyProGlyPro-GlyAsnProGlyAsnHisGlyGlyProGlnGlnGlnGlnGlnPro 331
 QY 377 TAAATTGACTGCTTGAAGACAGTATACCTCTCTCATCAGCTCAGACCCCTCTGA 318
 Db 332 -----ProGlnProProMet----- 336
 QY 317 CTTCGATGGGAGTACTATCTATCACTCCAGAGAGGCTCCGCTCTCCCTTGAA 258
 Db 337 -----AsnAsnGlyGlnMetGlyProProProLeuHis 347
 QY 257 GGGCTTCCCACTGCTCACTGTCCAGGTGTCTGTGTGTCACCAAGAGCCGAC 198
 Db 348 Ser-----ProLeuGlyAsn-----GlyPro 354
 QY 197 ACTGGGATGGCTCTCACAGA-----CATCCACAGCTCAGCCAGGTTGAACACTGAA 144
 Db 355 ThrGlyHisGlySerHisMetProGlyGlyProIleProGlyProGlyProGly 374

QY 143 GGAGCAGCTCCCGTGTGTGTAGATGCGCAGCAGCGTAGTCTTCTGGGGA 90
 Db 375 GlyLeuValGlyProGlyGlyIleSerProAlaHisGlyAsnAsnProGlyGly 392

Search completed: June 24, 2003, 13:57:50
 Job time: 155 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - protein search, using frame_plus.n2p model

Run on: June 24, 2003, 13:41:27 ; Search time 27 Seconds
(without alignments)
4292.032 Million cell updates/sec

Title: US-09-989-919-15

Perfect score: 2527
Sequence: 1 ggtctgcacgtctaccgga.....aaaaaaaaaagcggtc 1397

Scoring table: BLOSUM62
Xgapop 10.0, Xgapext 0.5
Ygapop 10.0, Ygapext 0.5
Fgapop 6.0, Fgapext 7.0
Delop 6.0, Delext 7.0

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 225784

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame+ n2p model -DRV=xlp
-O/cg2n2.1/USFto spool/US09989919/runat.24062003.102046.19173/app.query.fasta.1.1543
-DB=SwissProt_40 -QPM=fastran -SUFF=isp -MINMATCH=0.1 -LOOPEL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.coi -LIST=45
-DOCAIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=pcr -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000
-USER=US09989919 @CCN 1.1.31 @runat.24062003.102046.19173 -NCPU=6 -ICPU=3
-NO MMAP -LARGEOUTERY -NEG_SCORES=0 -WAIT -DSFBLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : SwissProt_40.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
C 1	126	4.9	1239	1 NME3 MOUSE	001098 mus musculu
C 2	119.5	4.7	1063	1 POLS RUBVR	P19725 rubella vit
C 3	119	4.6	1063	1 POLS RUBVR	P21480 rubella vit
C 4	119	4.6	1063	1 POLS RUBVR	P07566 rubella vit
C 5	118	4.6	897	1 CYR0 HUMAN	P32927 homo sapien
C 6	117	4.6	775	1 ICPO HSV1	P08393 herpes simp
C 7	117	4.6	1185	1 DRPL HUMAN	P54259 homo sapien
C 8	114	4.5	810	1 ILAR MOUSE	P16382 mus musculu
C 9	113	4.4	1233	1 NME3 HUMAN	Q14957 homo sapien
C 10	112.5	4.4	358	1 KLFI MOUSE	P46039 mus musculu
C 11	112.5	4.5	1964	1 NTC4 MOUSE	P31695 mus musculu
C 12	112	4.4	907	1 VGP3 EBV	P03200 Epstein-bar
C 13	112	4.4	992	1 POLS RUBVM	P08563 rubella vit
C 14	112	4.4	3149	1 TEGU EBV	P03186 Epstein-bar
C 15	111	4.3	707	1 SFPO HUMAN	P23246 homo sapien
C 16	109.5	4.3	1237	1 NME3 RAT	Q00961 rattus norv
C 17	109.5	4.3	1298	1 ICPI HSV1	P08392 herpes simp
C 18	108.5	4.2	978	1 PEX6 RAT	P54777 rattus norv

C 19	108.5	4.2	1362	1 BRD4 HUMAN	O60885 homo sapien
C 20	108	4.2	2167	1 SHK1 RAT	Q9w48 rattus norv
C 21	107	4.2	647	1 GAG SFV1	Q00071 simian foam
C 22	107	4.2	747	1 ELS BOVIN	P04985 bos taurus
C 23	107	4.2	1247	1 CTD2 MOUSE	O35927 mus musculu
C 24	107	4.2	2319	1 NTC3 RAT	Q9r172 rattus norv
C 25	107	4.2	3530	1 MY15 HUMAN	Q9ukn7 homo sapien
C 26	106.5	4.2	649	1 DRB3 HUMAN	Q16643 homo sapien
C 27	106	4.1	426	1 EXLP TOBAC	Q03211 nicotiana t
C 28	105	4.1	557	1 HNF8 HUMAN	P35680 homo sapien
C 29	105	4.2	717	1 PRD0 HUMAN	Q9h4q3 homo sapien
C 30	104.5	4.1	2142	1 BAT2 HUMAN	P48634 homo sapien
C 31	104.5	4.1	550	1 VGLE HSV1	P04488 herpes simp
C 32	104.5	4.1	780	1 YNL5 CAEEL	Q21955 caenorhabdi
C 33	104	4.1	660	1 YHL1 EBV	P03181 Epstein-bar
C 34	103.5	4.0	352	1 GLN2 PRAL	P20805 frankia aln
C 35	103.5	4.1	2003	1 NTC4 HUMAN	Q99466 homo sapien
C 36	101.5	4.0	1336	1 W146 HUMAN	Q9c0j8 homo sapien
C 37	100.5	3.9	1075	1 NFG3 HUMAN	Q12368 homo sapien
C 38	100	3.9	557	1 HNF8 RAT	P23699 rattus norv
C 39	100	4.0	1173	1 TSP1 XENLA	P35448 xenopus lae
C 40	100	3.9	2318	1 NTC3 MOUSE	O61982 mus musculu
C 41	100	3.9	3164	1 TEGU HSV1	P10220 herpes simp
C 42	99.5	3.9	4393	1 PG8M HUMAN	P98160 homo sapien
C 43	99	3.9	533	1 RXRB HUMAN	P28702 homo sapien
C 44	99	3.9	823	1 NUC1 NEUCR	P20824 neurospora
C 45	99	3.9	867	1 SSPO BOVIN	P98167 bos taurus

ALIGNMENTS

RESULT 1
NME3 MOUSE
ID NME3 MOUSE STANDARD; PRT; 1239 AA.
AC 001038;
DT 01-JUN-1994 (rel. 29, Created)
DT 01-FEB-1996 (rel. 33, Last sequence update)
DT 15-JUN-2002 (rel. 41, Last annotation update)
DE Glutamate (NMDA) receptor subunit epsilon 3 precursor (N-methyl
D-aspartate receptor subtype 2c) (NR2c) (NMDAR2c).
GN GRIN2C.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus;
OX NCBI_Taxid=10090;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=92310564; PubMed=1377365;
RA Kutsuwada T., Kashiwabuchi N., Mori H., Sakimura K., Kushiya E.,
RA Araki K., Meguro H., Maeki H., Kumanishi T., Arakawa M.,
RA Mishina M.,
RT "Molecular diversity of the NMDA receptor channel."
RL Nature 358:36-41 (1992).
RN [2]
RP REVISIONS.
RA Kashiwabuchi N.;
RL Submitted (SEP-1994) to the EMBL/GenBank/DDJ databases.
CC - FUNCTION: NMDA RECEPTOR SUBTYPE OF GLUTAMATE-GATED ION CHANNELS
CC - POSSESSES HIGH CALCIUM PERMEABILITY AND VOLTAGE-DEPENDENT
CC - SENSITIVITY TO MAGNESIUM AND IS MEDIATED BY GLYCINE.
CC - SUBUNIT: HETERODIMER OF AN EPSILON SUBUNIT AND A ZETA SUBUNIT.
CC - SUBCELLULAR LOCATION: Integral membrane protein.
CC - SIMILARITY: BELONGS TO THE LIGAND-GATED IONIC CHANNEL FAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; D10694; BAA01536.1; -

DR HSSP: P19491; 1GR2.
 DR MGD; MGI:95822; Grin2c.
 DR InterPro; IPR001320; Ion_glu_receptor.
 DR InterPro; IPR001311; SBP/glu_receptor.
 DR Pfam; PF00060; lig_chan; 1.
 DR Prodom; PD000500; Ion_glu_receptor; 1.
 DR SMART; SM00079; PDBE; 1.
 DR Receptor; Signal; Transmembrane; Postsynaptic membrane; Calcium;
 KW Ionic channel; Magnesium.
 FT CHAIN 1 19
 FT SIGNAL 20 1239
 FT DOMAIN 20- 553
 FT TRANSMEM 554 574
 FT TRANSMEM 597 617
 FT TRANSMEM 627 647
 FT TRANSMEM 815 835
 FT SITE 612 612
 FT CARBOHYD 70 70
 FT CARBOHYD 337 337
 FT CARBOHYD 438 438
 FT CARBOHYD 539 539
 SQ SEQUENCE 1239 AA; 135420 MW; 793E8E731E20C39 CRC64;
 Alignment Scores:
 Pred. No.: 0.037 Length: 1239
 Score: 126.00 Matches: 94
 Percent Similarity: 30.63% Conservative: 23
 Best Local Similarity: 24.61% Mismatches: 112
 Query Match: 4.92% Indels: 153
 Gaps: 24
 DB: 1
 US-09-989-919-15 (1-1397) x NME3_MOUSE (1-1239)
 QY 1196 AGAAGAGCCCGGCGCCCGCCAGAGCTGCTGCTCAACACACAGCTGCTCCCTGA 1137
 DB 927 ArgArgAlaProAlaPro-----ThrThrSerGlyPro----- 937
 QY 1136 CCCAGCTTGGCTGATGCTGAGGTGTCACCTCTGCCCTGACCCCTGGG- 1086
 DB 938 -----ArgSerCysThrProGlyPro-----ProGlyGlnProSer 949
 QY 1085 -----CTGGCTGGGAGTGTCCCTGTCTGGGAGGTAGAGTACCTCCCGAGTACGAAT 1029
 DB 950 ProSerGlyTyrPargProPro-----GlyGlyGlyValArg--ThrProLeuAlaArgArg 966
 QY 1028 CACCCACATCTGAGAGATCCCGCCGCAAG-----TCTCTGTCCAGACTTACGT 978
 DB 967 AlaProGlnProProAlaArgProAlaThrCysAlaGlySerProGlnProAspValSer 986
 QY 977 GAGAGAACACTCCAGATGATGTCCCGACGACCCCGCCAGCGCTGTGTGCATGTG 918
 DB 987 ArgAlaSerCysArgHisAlaTyrPaspAlaArg-----TyrProVal 1000
 QY 917 CAGACAGACAGAGCGGCTTACTTACATTTTCACACTATTTGAACAGAGTACGGGCT 858
 DB 1001 ArgValAlaGlyHisGlyGly-----SerHisLeuSerAlaSer 1012
 QY 857 GCCCAGGAAACCGCCAGGAGGAGGAGGAGAGCCCGCAGTGGGAGAGTACGCTGAGCTTG 798
 DB 1013 GlnArgArgAlaLeuProGlnArg-----SerLeu 1022
 QY 797 CTTACAGCCAC-----AGGGTTGAGAGGAGCAGTC 768
 DB 1023 LeuHisAlaHisCysHisTyrSerSerPheProArgAlaGlnArgSerGlyArgProPhe 1042
 QY 767 ATACCCCATATTCCTCGAGCTAGAGCCAGCA-----GTCCCATACTAGTGGC 720
 DB 1043 LeuPro-----LeuPheProGlnProProGlnProAspAspLeuProLeuGly 1059
 QY 719 TGGTAAACTCATTTTAACTTTGAGCTATTTGACAGCTTTGTCTGTCTGTCACATCTGTCT 660

DB 1060 ---ProGlnGlnLeuAlaArgArgGlnAlaLeuLeuArgAlaAlaTyrPala----- 1075
 QY 659 GTCTGGAGATGTAGCTTTGTCTGACAGGGGTTCCCAAG---TGACATGAGAGTCCCACT 603
 DB 1076 -----ArgGly- ProArgProArgHisAlaSerLeuProSe 1087
 QY 602 GCAAGTC-----ACTCCCTCCACAGGAGCCAGCAGC---TCTGAGATAGTAC 558
 DB 1087 rSerValAlaGlnAlaPheThrArgSerHisProLeuProAlaArgCysThrGlyHisAl 1107
 QY 557 AGCTGCTAGACCCGAGCCCTCAGATGTGTGACGACGAGCCCTTCACATATGTGCTTTGT 498
 DB 1107 aCysAlaCysProCysProGln----- 1114
 QY 497 GGGATCAGGAGCACTTGCTGCTCATTCAGTGTGAAAAAGACAGCTGCGCA----- 446
 DB 1115 -----SerArgProSerCysArgHisValAl 1123
 QY 445 -----CTGGGGGAGAGAGTGTAGAGTACAGAGGCTGAGT 408
 DB 1123 aglnThrGlnSerLeuArgLeuProSerTyrArgGlnAlaCysValGlnGlyValProAl 1143
 QY 407 TTTCTTTCTGGAGGCAAGACCATTTCTCTTAATTGACTGCTTGAAGACAGCTATACC 348
 DB 1143 aglyValAlaAla-----ThrTyrGlnThr 1151
 QY 347 TCTCTCTCAT-----CACTGACAGACCCCTCTGACTTCGATGGGGGATCATAT 297
 DB 1151 aArgGlnHisValCysLeuHisThrHisThrHisLeuProPheCysTyrGlyThrValCys 1171
 QY 296 CTATCAACCTCCAGAGAGGTCTGCCATCTCCCTTGAAGAGGCTTCTCCAGCTGCTAC 237
 DB 1171 aArgHisProProProCysSerSerHisSerProThr----- 1183
 QY 236 CTGTCCAGGTGTCTGTGTTGT-----GACCAAGAGCCCGACACTGGG 192
 DB 1184 -----LeuIleGlyThrTyrGlnProProSerHisArgGlyArgThrLeuGln 1199
 QY 191 CANG 188
 DB 1199 yLeu 1200
 RESULT 2
 POS RUBVR STANDARD: PRT: 1063 AA.
 ID POS RUBVR
 AC P19725; Q86370; Q86371; Q86372;
 DT 01-FEB-1991 (Rel. 17, Created)
 DT 01-FEB-1991 (Rel. 17, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Structural polyprotein [contains: Nucleocapsid protein C; Membrane glycoprotein E2; Membrane glycoprotein E1].
 OS Rubella virus (vaccine strain RA27/3).
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Togaviridae;
 OC Rubdivirus.
 OX NCB1 Taxid=11044;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=89296505; PubMed=2740235;
 RA Nakhasi H.L., Thomas D., Zheng D., Liu T.Y.;
 RT "Nucleocapsid sequence of capsid, E2 and E1 protein genes of Rubella virus vaccine strain RA27/3."
 RL Nucleic Acids Res. 17:4393-4394(1989).
 CC - FUNCTION: GLYCOPROTEIN E1 CONTAINS THE VIRAL HEMAGGLUTININ ACTIVITY.
 CC - SUBUNIT: MULTIPLE COPIES OF THE C PROTEIN COMPRISE THE NUCLEOCAPSID.
 CC - DISEASE: RUBELLA INFECTION DURING THE FIRST TRIMESTER OF PREGNANCY RESULTS IN DEATH OR MULTISYSTEM BIRTH DEFECTS & MICROCEPHALY.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way

CC -1- FUNCTION: GLYCOPROTEIN E1 CONTAINS THE VIRAL HEMAGGLUTININ
 CC ACTIVITY.
 CC -1- SUBUNIT: MULTIPLE COPIES OF THE C PROTEIN COMPRISE THE
 CC NUCLEOCAPSID.
 CC -1- DISEASE: RUBELLA INFECTION DURING THE FIRST TRIMESTER OF PREGNANCY
 CC RESULTS IN DEATH OR MULTISYSTEM BIRTH DEFECTS & MICROCEPHALY.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL: M30776; AAA47421.1;
 CC PIR: J00087; GNMV77.
 DR Polypeptide; Glycoprotein; Transmembrane; Signal.
 KM CHAIN 1 300 NUCLEOCAPSID PROTEIN C.
 FT CHAIN 301 562 MEMBRANE GLYCOPROTEIN E2.
 FT SIGNAL 563 582
 FT CHAIN 583 1063
 FT TRANSMEM 278 300
 FT TRANSMEM 517 535
 FT TRANSMEM 538 555
 FT CARBOHYD 353 353
 FT CARBOHYD 371 371
 FT CARBOHYD 429 429
 FT CARBOHYD 658 658
 FT CARBOHYD 759 759
 FT CARBOHYD 791 791
 SQ SEQUENCE 1063 AA; 114713 MW; 18E77B28B9A667 CRC64;

Alignment Scores:
 Pred. No.: 0.128 Length: 1063
 Score: 119.00 Matches: 98
 Percent Similarity: 32.46% Conservative: 39
 Best Local Similarity: 23.22% Mismatches: 127
 Query Match: 4.65% Indels: 159
 DB: 1 Gaps: 22

US-09-989-919-15 (1-1397) x: POLS_RUBVH (1-1063)

QY 1184 GCCCCACCCCGCCCGCCAGCTCTGCTCAAC-----ACACTGCTCCCTGACCC 1134
 DB 77 AlarProProProProGluGluArgGlnGlnSerArgSerGlnThrProAlaProLysPro 96
 QY 1133 CAGCTTGGGCTCAGCTGAGGTGTGCACTCTGCTGACCCCTGGGGTGGCTGGAG 1074
 DB 97 SerArg-----AlaPro-----ProGlnGlnProGln----- 105
 QY 1073 TGTCCCTGTGTGGAGGTAGAGATAGCTCCAGGTACG-----AAT 1029
 DB 106 TGTCCCTGTGTGGAGGTAGAGATAGCTCCAGGTACG-----AAT 1029
 QY 1028 CACCCACATCTGTGAGCATCCCGCCAGCTCTCTGTCCAGACTTAAGTGAAGA--- 972
 DB 117 AlarProArgProGluLeuGlyProProThrThrProPheGlnAlaAlaValAlaAlaArgGly 136
 QY 971 -----ACACTTCCAGATGATTTGCCACGACGACCCCGCCAGCTCTGT----- 926
 DB 137 LeuArgProProLeuHisAspPro-----AspThrGlnAlaAlaProThrGlnAlaCysValTh 155
 QY 925 -----GGCATGTGACGACGACGACGAGGGGCTTAACTCACTTCTCACTATTGAACAC 870
 DB 155 rSerTrpLeuTrpSerGlnGlyGlnGlyAlaValAlaPheTrpArgValAspLeuHisPheTh 175
 QY 869 AGTATGGGGCTGCCCGACGGAACGCTCCAGGGA----- 836
 DB 175 rAsnLeuGlyThrProProLeuAspGlyArgTrpAspProAlaLeuMetClyrAs 195
 QY 835 -----GGGGACACGCCCCACGATGGGAGAGACTGACCTTGTACAGCCCAAGG 783

DB 195 nProCysGlyProGlnProPro-----AlaHisValValArgAlaTrpAsnGlnProAlaG 214
 QY 782 GTTGAGAGGCGATCATACCCCATGATTCCTGACATAGAGCCAGACGATCCCATAGCTA 723
 DB 214 yAspValArg-----GlyValTrpGlyLeuGlyGlnArgTrpTrpAla- 228
 QY 722 GGGCTGTAACTCATATTTAACAT-----TGSCATTAATTCACAGTTGTGCC 675
 DB 229 -----GlnGlnAspPheArgValGlyGlyThrArgTrpHisArgLeuLeuMetP 246
 QY 674 TGTGCATCATCTGTCTGTGTGGAGATTTAGCTTTGTCTGCAGAGGTTCCAGGTGACATGC 615
 DB 246 roval-----ArgGlyLeuAspGlyAsp----- 253
 QY 614 AGTCTGCCAGTGCATGATCATCTCCCTCCACAGGACGAGCCAGCTGTGAGATAG-TCAG 556
 DB 254 -----ThrAlaProLeuProProHisThrThrGlnArgGlyLeuThrArgSerA 270
 QY 555 CTGTGACGCG-AGCCCTCAGATAGGTGACAGGAGCCCTTCACATATGTGCTTGT 499
 DB 270 lArgArgProTrpArgGlyLeuArgPheGlyAlaProGlnAlaPheLeuAlaGlyLeuLeu 290
 QY 498 TGGATTCAGGACCACTTGGCTTCATCCAGTCTTGAAAAAGACCACTGCGGAGCTGGGG 439
 DB 290 euAlaAlaValAla-ValGlyThrAlaArgAlaGlyLeuGlnProArgAlaAspMet--- 308
 QY 438 GCGAAGGTGAGAGTAAACAGGCGTGTGATGTTCTCTTCTGAGAGCAAGCATTTCC 379
 DB 308 ----- 308
 QY 378 CTAATTGACGTCTTGAAGACAGATATACCTCTCTCATCATCACTGACACCCCTCTG 319
 DB 309 -----AlaAlaProPromet 313
 QY 318 ACTTCGATGGGGATCACTATCTATCAACTCCAGAGAGCTCTGCAATCTCCTTGA 259
 DB 314 -----Pro-ProGlnPro----- 317
 QY 258 AGGCTTTCCTCCAGTGGCTCACCTGTCCAGAGGTGTGTTGGTGACCAAAAGCCGCA 199
 DB 318 -----ProArgAlaHisGlyGlnHisIstyGlnHisIshHisIshGlnLeuProPh 334
 QY 198 CACTGGGATGAGCTCTTCACGACATCCACAGCTCAGCCAGGTGAACCTGAAGAGAG 139
 DB 334 eleuGlyHisAspGlyHisHis-----GlyG 343
 QY 138 CAGCTCCCTGTGTGTGATGGCCAGACGCGGTAGT-----CT 100
 DB 343 yThrLeuArgValGlyGlnHisHisIshArgAsnAlaSerValLeuProGlyHisIshIsh 363
 QY 99 TCTGTGGGATGTGTGTGTCTGGA-----TACAGTGTGACTCGGTACTGC 54
 DB 363 uGlnGlyGlyTrpGlyCysTyTrpAsnLeuSerAspTrpHisGlnGlyThrHisValCys 382

RESULT 4
 POLS_RUBVT STANDARD; PRT; 1063 AA.
 AC P07566;
 DT 01-APR-1988 (Rel. 07, Created)
 DT 01-AUG-1988 (Rel. 08, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Structural polyprotein [contains: Nucleocapsid protein C; Membrane glycoprotein E2; Membrane glycoprotein E1].
 OS Rubella virus (strain Therien).
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Togaviridae;
 OC Rubivirus.
 OX NCBI_TaxID=11045;
 RN [1].
 RP SEQUENCE FROM N.A.
 RX MEDLINE=90281585; PubMed=2353453;
 RA Dominguez G., Wang C.Y., Frey T.K.;
 RT "Sequence of the genome RNA of rubella virus: evidence for genetic rearrangement during togavirus evolution.";

QY 717 GTHAACTACATTAACTTTGACATTGACAGCTTTGCTGCTGACATCTGCTGT 658
 Db 641 ----- 641
 QY 657 CTGGAGTGTAGCTTTCTTCGAGGGTTCAGAGTACATGACGTCTCCAGTGAAG 598
 Db 642 -----LeuAlaGlnAlaMetGlyProGlyGln-----AlaValGlnValGlnA 656
 QY 597 TCACTCCCTCCACAGCAAGCCAGCTGCTGAGATGATGAGTGTGTCAGGACCCCTC 538
 Db 656 rgarProSerGingly-----AlaAlaGlySerProSerLeug 669
 QY 537 AGATAGCTGACGAGAGCCCTTACATATGTGCTTTGTTGGATGAGGACCACTTGC 478
 Db 669 luserGlyGly-----GlyProAlaProProAla 679
 QY 477 TCCATCCAGCTTGAAGAAACAGCAGCTCCGACCTGGGGGAGAGGTGAGTAAAC 418
 Db 679 euGlyPro-----ArgValGlyGlyGlnAspGlnLysAspSer 692
 QY 417 GGCTGATGTTCTCTTCTTCGAGCAAGACATTCCTCAATTTGACTGCTTGAAGA 358
 Db 692 roValAlaIleProMetSerSerGlyAspHrGlnAspProGlyValAlaSerGlyTyv 712
 QY 357 CACGTATACCTCTCTCTCATCAGCTGACACCCCTCTGACTTCCGATGGGGATCACTA 298
 Db 712 alser-SerAlaAspLeuVal-----PheThrProAsnSerGlyAlaSer 726
 QY 297 TCTATCAACTC-----CAGAGAGTCTCTGCCATCTCCCTT 262
 Db 727 SerValSerLeuValProSerLeuglyLeuProSerAspGlnThrProSerLeuCysPro 746
 QY 261 GGAAGGCTTCT-----CCCAGCTCAGCTGTCAGGTG----- 226
 Db 747 GlyLeuAlaSerGlyProProGlyAlaProGlyProValIlySerGlyPheGlyGlyTyv 766
 QY 225 GTCTGTGTGTGACCAAGAGCCGACACTGGGATGCTCTCACAGATCCACAGCC 166
 Db 767 ValGlnLeuProProIleGlnGlyArg-SerProArgSerProArgAsnAsnProValPr 786
 QY 165 T-----CAGCCAGGTTGAACACTGAAGAGAGGAGAGCTCCG 130
 Db 786 oProGlnAlaLysSerProValLeuAsnProGly-----Gluar 799
 QY 129 TGTGTGATGATGACCAAGAGCCGATGCTCTCTGGGGGATGATGCTGTGGATACAC 70
 Db 799 gProAlaAspValSerProThrSerProGlnProGlyLeuLeuValLeuGlnGlnVa 819
 QY 69 TGGTACTCGTACTGCTGCTTGGCG 45
 Db 819 lGlyAsp---TyrCysPheLeuPro 826

RESULT 6
 ICP0_HSV11
 ID ICP0_HSV11 STANDARD; PRT; 775 AA.
 AC P08393;
 DT 01-AUG-1988 (Rel. 08, Created)
 DT 01-AUG-1988 (Rel. 08, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Trans-acting transcriptional protein ICP0 (Immediate-early protein IE110) (VNM110) (Alpha-0 protein).
 GN ICP0 OR IE110.
 OS Herpes simplex virus (type 1 / strain 17).
 OC Viruses: dsDNA viruses, no RNA stage; Herpesviridae;
 OC Alphaherpesvirinae; Simplexvirus.
 NC NCBL_TaxID=10299;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=88274327; PubMed=2839594;
 RA McGeoch D.J., Dalrymple M.A., Davison A.J., Dolan A., Frame M.C.,
 RA McNab D., Perry L.J., Scott J.E., Taylor P.;
 RT "The complete DNA sequence of the long unique region in the genome of
 herpes simplex virus type 1.";

RL J. Gen. Virol. 69:1531-1574(1988).
 RN [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=87059760; PubMed=3023529;
 RA Perry L.J., Rixon F.J., Everett R.D., Frame M.C., McGeoch D.J.;
 RT "Characterization of the IE110 gene of herpes simplex virus type 1.";
 RL J. Gen. Virol. 67:2365-2380(1986).
 RN [3]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=89036163; PubMed=2846760;
 RA Perry L.J., McGeoch D.J.;
 RT "The DNA sequences of the long repeat region and adjoining parts of
 the long unique region in the genome of herpes simplex virus type
 1.";
 RL J. Gen. Virol. 69:2831-2846(1988).
 CC -1- FUNCTION: ICP0 IS A GENERAL TRANSDUCTION FACTOR OF ALL THREE CLASSES OF
 HSV GENES WHICH FUNCTIONS SYNERGISTICALLY WITH ICP4. OFTEN TERMED
 AS A PROMISCUOUS TRANSDUCTION FACTOR. MAY PLAY A ROLE IN THE BALANCE
 BETWEEN THE LATENT AND LYtic STATES, REACTIVATING LATENT HSV. ICP0
 MIGRATES TO THE ND10 NUCLEAR BODIES IN THE EARLY TIMES OF
 INFECTION AND DISRUPTS THEM. BINDS TO HAUSP AND MAY MODIFY ITS
 DEUBIQUITINATING SUBSTRATE SPECIFICITY OR ACTIVITY ON VIRAL OR
 CELLULAR TARGETS, LEADING TO AN INCREASED OR DECREASED STABILITY
 OF THESE PROTEINS.
 CC -1- SIMILARITY: CONTAINS 1 RING-TYPE ZINC FINGER.
 CC -1- SIMILARITY: TO OTHER HERPESVIRUSES ICP0 PROTEIN.
 CC CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).
 CC
 DR EMBL; X14112; CAA32336.1; -
 DR EMBL; X14112; CAA32293.1; -
 DR EMBL; X04614; CAA28285.1; -
 DR PIR; A29152; EDBE11.
 DR InterPro; IPR001841; Znf_Ring.
 DR Pfam; PF00097; zf-CHC4; 1.
 DR SMART; SM00184; RING; 1.
 DR PROSITE; PS00518; ZF_RING_1; 1.
 DR PROSITE; PS50089; ZF_RING_2; 1.
 KW Transcription regulation; Trans-acting factor; Activator; Zinc-finger;
 KW DNA-binding; Early protein.
 FT ZN FING 116 157 RING-TYPE.
 FT ZN FING 233 243 POLY-ASP.
 FT DOMAIN 305 308 POLY-GLY.
 FT DOMAIN 558 568 POLY-SER.
 SQ SEQUENCE 775 AA; 78456 MW; DF38A1C539DAB15C CRC64;

Alignment Scores:
 Pred. No.: 0.176 Length: 775
 Score: 117.00 Matches: 47
 Percent Similarity: 36.00% Conservative: 7
 Best local Similarity: 31.33% Mismatches: 61
 Query Match: 4.57% Indels: 35
 DB: 1 Gaps: 8

US-09-989-919-15 (1-1397) x ICP0_HSV11 (1-775)
 QY 1188 CCGGCGCCACCCCGCCAGCTGCTCAACACACCTGCTCCCTGACCCGACGTC 1129
 Db 381 ProGlySerAlaProArgProGlyProProAlaSerAlaAlaSer---GlyProAla 399
 QY 1128 TTGGCTCATGCTGAGGTG-----TGCACCTCTGCTGACCCCTGGGCTGGCTGGCA 1075
 Db 400 ArgProArgAlaAlaValAlaProCysValAlaGlnAlaProPro----- 414
 QY 1074 GTGTCCCTGTGCGGGAGTCAAGATAGCTCCCGAGTACAGATCAACCACTCTGG 1015
 Db 415 -----GlyProGlyProArgAlaProAla-ProG 424

Db 448 SerTrpAlaCysLeuProMetGlyProSerGlnGlnAlaThrCysGlnAlaThrGlnGln 467
 QY 1025 CCA---CATCTCTGA----- 1014
 Db 468 ProSerHisProGlyProLeuSerGlySerProAlaGlnSerAlaProThrLeuAlaCys 487
 QY 1013 -----GCA 1011
 Db 488 ThrGlnValProLeuValLeuAlaAspAsnProAlaTyrArgSerPheSerAspCysCys 507
 QY 1010 TCCCGCGCAAGTCTCTGTCAGACCTAGTACAGACCTCCAGCATGATGATG--- 954
 Db 508 SerProAlaProAsnProGlyGlnLeuAlaProGlnGlnGlnAlaAspHisLeuGln 527
 QY 953 -----TCCCGACGACGCGCCGCGCTG---CTGGGACATG 918
 Db 528 GluGlnGluProSerProAlaAspProHisSerSerGlyProProMetGlnProVal 547
 QY 917 CAGCAGACGACGAGGGGCTTTAGCTCATCTTCACACTTGAACACATGATGGGGCT 858
 Db 548 GluSerTrpGlnGln-----IleLeuHisMetSerValLeuGlnHis-----GlyAla 563
 QY 857 GCCCAGCAGACCGTCCGACGG----- 837
 Db 564 AlaAlaGlySerThrProAlaProAlaGlyGlyTyrGlnGlnGlnValGlnAlaValys 583
 QY 836 -----AGGGGACAGCCCCAGTGGGAGACTAGCTG 804
 Db 584 GlnGlyAlaAlaGlnAspProGlyValProGlyValArgProSerGlyAspProGlyTyr 603
 QY 803 -----AGCTTGCTTACGCCCCACAGGTT----- 780
 Db 604 LysAlaPheSerSerLeuSerSerAsnGlyIleArgGlyAspThrAlaAlaAlaGly 623
 QY 779 -----GAGAGGAGCATCATACCCCATGATCTGACATGAGAGCA----- 728
 Db 624 ThrAspArgGlyHisGlyGlyTyrLysPro---PheGlnAsnProValProAsnGlnSer 642
 QY 737 -----GAGTCCCATAGTACGCTGTAACTCACTATTTACATTTGGCATTTATGCA 684
 Db 643 ProSerValPro-----LeuPheThrPheGlyLeuAspThr 655
 QY 683 CGTTGCTGTCGTCACATCTGTC-----TGCTGCGA 652
 Db 656 GluLeuSerProSerProLeuAsnSerAspProLysSerProProGluCysLeuGly 675
 QY 651 TGTTAGCTTTGTCGACGAGGCTTCCAGTACATGACATGCTGCCAGTGCATGCACTC 592
 Db 676 LeuGlnLeuGlyLeuLysGly-----GlyAspTrp-----ValLysAlaPro 689
 QY 591 CCTCCACAG-----GACAGCCGACCTGCTAGAGTACTACGCTGTCAGCCGAC 541
 Db 690 ProProAlaAspGlnValProLysProPheGlyAspAsp---LeuGlyPheGlyIleValTy 709
 QY 540 CTCAGATAGTCAGACGACGAGGCTTCATATGTCGTTGTGGATCAGAGGACCACT 481
 Db 709 rSerSerLeuThrCys-----HisLeuCysGly-----HisLeuGlnHis-- 723
 QY 480 GGTTCATCTTCAGTTGAAAAAGACCAAGCTCCGACCTGGGGGACAGGCTGAGAGGTTAA 421
 Db 724 -----HisSerGlnGlnGlnGlyGlnSerProIle-----ValAlaSerPr 738
 QY 420 CAGGCGTGGATGTTCTCTTCTTGAGGACAGACCATTCCTCAATTTAGCTGCTTGA 361
 Db 738 oglyCysGlyCys-----CysTyrAs 745
 QY 360 AGACAGCATACCTCTCTCATACCTCAGACCCCTCTGACTTCGATGGAGGAGATCA 301
 Db 745 pasparSerProSerLeuGlySer----- 753
 QY 300 CTATCTATCAACCTCCAGAGAGGCTCTGCACTTCCTTGAAGGGCTTCTCCACATGGC 241

Db 754 -----LeuSerGlyAlaLeuGlnSerCysPro-----GluGlyIleProProGlnAl 769
 QY 240 TCACCTGTCACAGT 181
 Db 769 AsnLeuMetSerAla-----ProLysThrProSerAsnLeuSerGlyGlnG 785
 QY 180 CAGACATCCACAGCCTCAGCCA 159
 Db 785 ylyGlyGlyProGlyHisSerPro 792
 RESULT 9
 NM3 HUMAN STANDARD; PRT; 1233 AA.
 ID NM3 HUMAN
 AC 014957;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Glutamate (NMDA) receptor subunit epsilon 3 precursor (N-methyl
 D-aspartate receptor subtype 2c) (NR2C) (NMDAR2C).
 GN GRIN2C.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
 OX NCBI_TaxID=9606;
 RN (1)
 RP SEQUENCE FROM N.A.
 RC Tissue=Brain;
 RX MEDLINE=97189248; PubMed=9037519;
 RA Lin Y.J., Boveito S., Carver J.M., Giordano T.;
 RT "Cloning of the cDNA for the human NMDA receptor NR2C subunit and its
 expression in the central nervous system and periphery."
 RL Brain Res. Mol. Brain Res. 43:57-64(1996).
 CC - FUNCTION: NMDA RECEPTOR SUBTYPE OF GLUTAMATE-GATED ION CHANNELS
 POSSESSES HIGH CALCIUM PERMEABILITY AND VOLTAGE-DEPENDENT
 SENSITIVITY TO MAGNESIUM AND IS MEDIATED BY GLYCINE.
 CC - SUBUNIT: HETERODIMER OF AN EPSILON SUBUNIT AND A ZETA SUBUNIT.
 CC - SUBCELLULAR LOCATION: Integral membrane protein.
 CC - TISSUE SPECIFICITY: MAINLY IN BRAIN WITH PREDOMINANT EXPRESSION IS
 IN THE CEREBELLUM, ALSO PRESENT IN THE HIPPOCAMPUS, AMYGDALA,
 CAUDATE NUCLEUS, CORPUS CALLOSUM, SUBTHALAMIC NUCLEI AND THALAMUS.
 CC DETECTED IN THE HEART, SKELETAL MUSCLE AND PANCREAS.
 CC - SIMILARITY: BELONGS TO THE LIGAND-GATED IONIC CHANNEL FAMILY.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).
 CC EMBL: L76224; AAA88096.1;
 DR HSSP; P19491; IGR2.
 DR Genew; HGNC:4587; GRIN2C.
 DR MIM; 138254;
 DR InterPro; IPR001320; Ion_glu_receptor.
 DR InterPro; IPR001311; SBP/glu_receptor.
 DR Pfam; PF00060; lig_chan; 1.
 DR ProDom; PD000500; Ion_glu_receptor; 1.
 DR SMART; SM00079; PDBE; 1.
 KW Receptor; Signal; Transmembrane; Postsynaptic membrane; Calcium;
 KM Ionic channel; Magnesium.
 FT SIGNAL 1 19
 FT CHAIN 20 1233
 FT TRANSMEM 554 574
 FT TRANSMEM 597 617
 FT TRANSMEM 627 647
 FT TRANSMEM 815 835
 FT SITE 612 612
 FT CARBOHYD 70 70
 FT CARBOHYD 337 337
 1 (POTENTIAL).
 2 (POTENTIAL).
 3 (POTENTIAL).
 4 (POTENTIAL).
 FUNCTIONAL DETERMINANT OF NMDA
 RECEPTORS (BY SIMILARITY).
 N-LINKED (GLCNAC...) (POTENTIAL).
 N-LINKED (GLCNAC...) (POTENTIAL).

DR Pfam; PF00096; zf-C2H2; 3.
 DR ProDom; PD000003; Znf_C2H2; 2.
 DR SMART; SM00355; Znf_C2H2; 3.
 DR PROSITE; PS00028; ZINC_FINGER_C2H2_1; 3.
 DR PROSITE; PS00157; ZINC_FINGER_C2H2_2; 3.
 KW Transcription regulation; Activator; Zinc-finger; Metal-binding;
 KW DNA-binding; Nuclear protein; Repeat.
 FT DOMAIN 1 269
 FT DOMAIN 58 76 ASP/GLU-RICH (ACIDIC).
 FT DOMAIN 275 357 ZINC FINGERS.
 FT ZN_FING 275 299 C2H2-TYPE.
 FT ZN_FING 305 329 C2H2-TYPE.
 FT ZN_FING 335 357 C2H2-TYPE.
 FT ZN_FING 345 357 C -> Y (IN REF. 3).
 FT ZN_FING 345 357 C -> T (IN REF. 3).
 FT ZN_FING 345 357 G -> P (IN REF. 3).
 FT ZN_FING 345 357 S -> T (IN REF. 3).
 FT CONFLICT 116 116
 FT CONFLICT 223 223
 FT CONFLICT 287 287
 SQ SEQUENCE 358 AA; 3757 MW; 6FD59F294D845D1 CRC64;

Alignment Scores:

Pred. No.: 0.359 Length: 358
 Score: 112.50 Matches: 80
 Percent Similarity: 32.23% Conservative: 27
 Best Local Similarity: 24.10% Mismatches: 97
 Query Match: 4.39% Indels: 128
 Gaps: 18

US-09-989-919-15 (1-1397) x KLF1_MOUSE (1-358)

QY 1179 ACCCCGCCCCAGGT-----CTGGCTCAAAACACACCTGCTCCCTGACCCAGTC 1129
 DB 80 ThrasmhepProglYserGlyserProglYThserYThrYsalalaProser 99
 QY 1128 TTGGCTCATGTGAGTGTGACCTCTGCCCTGACCCCTGGGCTGGAGATGTC 1069
 DB 100 ValAlPro-----ValAlaGlnPheGluProProGluSerLeuAlaTy 115
 QY 1068 CCTGCGGAGGTGAGATGCTCCCA---GTAACAAATACCCACATCTGTAGC 1012
 DB 116 AlaGlyGlyProGlyLeuValThGlyProLeuYserGluGluThrSerTrpAla 135
 QY 1011 ATCCCGCCGAGCTCTGTCAGACCTAGTGAAGAACAACCTCCAGATGATGTC 952
 DB 136 HisPro-----ThrProArg----- 140
 QY 951 CCCAGCAGCCCGCCCTGCTGTGGCATGTGACAGCAGCAGGAGGCTTAACT 892
 DB 141 -----ProProAlaProGluProPheValAlaProAlaLeuAlaProGlyLeuAla 157
 QY 891 ACATTTCACTATTGAACACAGTGTGGGCTGCCAGGAAACCGTCCAGG----- 838
 DB 158 -----ProYsalGlnProSerTySer 165
 QY 837 -----GAGGGCAGACCCCGAGTGGGA----- 814
 DB 166 AsperatgaIaglySerValGlyGlyPheheProaGalaGlyLeuAlaValProAla 185
 QY 813 -----GACTAGCTGAGCTTGTCTTACAGCCCAAG 784
 DB 186 AlaProGlyAlaProTyGlyLeuLeuSerGlyTyProAlaLeuTyProAlaProGln 205
 QY 783 GGTGAGAGGAGCAGTATACCCCATGATCTCTGATCAGTGAAGCAGAGCCCATAGCT 724
 DB 206 ---TyGlyGly---HisPheGlnLeuPheargGlyLeuAlaValProSerAlaGlyGly 223
 QY 723 AGGCTGTAACACTACATTGAATTGACATTATGACGTTGTCCTGTCATCT 664
 DB 224 ThrAlaProProSerPheLeuasn----- 231
 QY 663 GTCTGTCTGGAGTGTAGCTTGTCTGACGAGGAGTCCAGAGTACATGACAGTCTGCCAG 604
 DB 232 ---CysLeuGly-----ProGly-ThrValAlaThrGlu 242

QY 603 TGCAAGTCACTCCCTCCACAGCAAGCCGAGTGTGAGATGATGACCTGTGACCCGA 544
 DB 242 uGlyAlaThrAlaIleAlaGlyAspAlaGlyLeu-----SerProGlyThrAl 258
 QY 543 GCCCTCAGATAGTGTGACGAGGCTT-----CACATATGTGG 505
 DB 258 AProProYsalatSerArgatGlnLeuAlaProYsalArgGlnAlaAlaThrCysG 278
 QY 504 TCTTTTGGG-----ATCCAGGAGCCACTGGCTCC 475
 DB 278 YHisGlyGlyCysGlyYserTySerYserYserSerHisLeuYsalAlaHisLeuArgTh 298
 QY 474 ATCCAGTCTGTAAGAAGACACAGCTGCGAGCTGG---GGGCGAGAGTGAAG----- 426
 DB 298 HisThrGlyGlyYserProTyAlaCysSerTrpAspGlyCysAspTrpAlaAlaAr 318
 QY 425 -----GTAACAGGCTGTGATGTTTC-- 405
 DB 318 gSerAspGluLeuThrArgHisTyArgGlyHisThrGlyHisArgProPheCysG 338
 QY 404 -----TCTTTCTGAGGAGCAGACCAT 384
 DB 338 YLeuCysProArgAlaPheSerArgSerAspHis 349
 RESULT 11
 NTG4_MOUSE
 ID NTG4_MOUSE STANDARD; PRT: 1964 AA.
 AC P31695; Q62389; Q62390; Q35442; Q9R1W9; Q88314; Q88315; Q9R1X0;
 DT 01-JUL-1993 (Rel. 26, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Neutrogenic locus notch homolog protein 4 precursor (Notch 4)
 DE [Contains: Transforming protein Int-3].
 OS NOTCH4 OR INT3 OR INT-3.
 GN Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 NCBI_Taxid=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA MEDLINE=92194507; PubMed=1312643;
 RX Robbins J., Blondel B.J., Gallahan D., Gallahan R.,
 RT "Mouse mammary tumor gene int-3: a member of the notch gene family
 RT transforms mammary epithelial cells";
 RL J. Virol. 66:2594-2599 (1992).
 RN [2]
 RP REVISIONS, SEQUENCE FROM N.A.
 RA MEDLINE=97294599; PubMed=9150355;
 RX Gallahan D., Callahan R.,
 RT "The mouse mammary tumor associated gene INT3 is a unique member of
 RT the NOTCH gene family (NOTCH4).";
 RL Oncogene 14:1883-1890 (1997).
 RN [3]
 RP SEQUENCE FROM N.A.
 RA TISSUE=lung, and Testis;
 RX MEDLINE=96281668; PubMed=8681805;
 RA Wtendaele H., Marazzi G., Wu G., Yan O., Sassoon D., Kitajewski J.,
 RT "Notch4/int-3, a mammary proto-oncogene, is an endotheial
 RT cell-specific mammalian Notch gene";
 RL Development 122:2251-2259 (1996).
 RN [4]
 RP SEQUENCE FROM N.A.
 RA Rowen L., Mahairas G., Oin S., Ahern M.E., Dankers C., Lasky S.,
 RA Loretz C., Schmidt S., Tipton S., Traicoff R., Zackrone K., Hood L.,
 RT "Sequence of the mouse major histocompatibility locus class III
 RT region";
 RL Submitted (OCT-1997) to the EMBL/GenBank/DBJ databases.
 RN [5]
 RP SEQUENCE OF 1436-1600 FROM N.A.
 RX MEDLINE=99252212; PubMed=102333982;
 RA Lee J.-S., Haruna T., Ishimoto A., Honjo T., Yanagawa S.-I.,
 RT "Intracasternal type A particle-mediated activation of the Notch4/int3
 RT gene in a mouse mammary tumor: generation of truncated Notch4/int3

FT	REPEAT	1168	1208	LIN/NOTCH 1
FT	REPEAT	1209	1242	LIN/NOTCH 2
FT	REPEAT	1243	1282	LIN/NOTCH 3
FT	REPEAT	1628	1657	ANK 1
FT	REPEAT	1661	1691	ANK 2
FT	REPEAT	1695	1724	ANK 3
FT	REPEAT	1728	1757	ANK 4

Alignment Scores:	
Pred. No.:	0.441
Score:	112.50
Percent Similarity:	27.12%
Best Local Similarity:	21.22%
Query Match:	4.45%
DB:	1
Length:	1966
Matches:	130
Conservative:	36
Mismatches:	161
Indels:	34
Gaps:	28

US-09-989-919-15 (1-1397) X NTC4_MOUSE (1-1964)

QY	6	GGACCTGTACCGAGGGGGAGATATGTGCAGAACTGCA-----	44
Db	903	Cys11leasprh1glyserSerTyRheCyatgCysProProGlyRheGlnGlyIysIeu	922
QY	45	-----CGCAAGACGACGTA	59
Db	923	CysGlnAspAsnMeAsnProCysGlyIProAsnProCysHisHisGlySerThrCysVal	942
QY	60	CCGAGTACCACT-----GTATCCCAAGACACCAATCCCCAGAAAGACTACCGCTGC-	112
Db	943	ProGlnProSerGlyTyValCysGlnCysValaProGly--TyGlnGlyGlnAsnCys	962
QY	113	-----TGCCATCTTACCAACACCGGAGCTGCCTCC---	143
Db	962	erLysValIleuGlnAlaCysGlnIleerGlnProCysHisIaenHisGlyThrCys-ThrSer	981
QY	144	-----TTTCACTGTCA-----ACCTGGCTGAGGCTGT---	171
Db	982	ArgProGlyGlyIrrhHisCysAlaCysArProProGlyRheValGlyIleuArg-CysGlnG	1001
QY	172	-----GGATCTCTGTGAGAGCAATGGCCAGTGTGGGGC-----	204
Db	1001	YAspValAspGluCysLeuAspArgProCysHisIProSerGlyThrAlaAlaCysHisSe	1021
QY	204	-----	204
Db	1021	rIeuAlaAsnIaIreTyrcysGlnCysLeuProGlyHisThrGlyGlnArgCysGluVa	1041
QY	205	-----CTTTGTGTCAACCAACGAA-----	224
Db	1041	IglIueArAspIeuCysGlnSerGlnProCysSerAsnGlyIySerCysGlnIleThr	1061
QY	225	-----CCACCTGGACAGGTGAGCCATGGGAGAGCCCTTCAAGAGG-----AG	268
Db	1061	rGlyProProProGlyIrrhThrCysHisCysProIyGlyIrrhGlnGlyProThrCysSe	1081
QY	269	ATGGCAGACACTCTCTGGAGGTTGATAGTATGCCCCATCGGAAGTCAGGGGGT	328
Db	1081	rHisIyAlaIeuSerCysGly-----IleHisIcCysHisAsn-----	1094
QY	329	GCTGAGGTGATGACGACAGGATATCGTGTCTTCAAGCACTCAATTAGGAGGAATGT	368
Db	1095	-----GlyGlyIeuCys	1098
QY	389	CTTGCTCCAGAAAGAAACAATCCACAGCCCTGTATCTCTCACCTC-----	434
Db	1098	SleuProSerProIySProGlySerProProIeuCysAlaCysIeuSerGlyRheGlnGly	1118
QY	435	-----TGCCCCCAGCTGGGACGCTGCTTTTTCAAAGACTGG	472
Db	1118	yProArCySleuThrProProAlaIrrProGly-----CysGln	1131
QY	473	ATGGAGCCAAAGTGTCCCTGCATGCC-----AACAAAGCAATATGTGAAGGGCTCTGG	526
Db	1131	y-----ProIerProCysIeuHisenGlyThrCysThrGlnThrProGly	1147

OY		527	CTGACSTATTGTGAGGGCTCGGGCTGTACAAGCTGAATCTCTAGACAGCTTGGGCTTCCTGT	586
Dd		1147	Y-----LeuGlyAsnPro--	115
OY		587	GGAGGAGATGACTTTGCACCTGGACACACTCATGTCACTGGAAACCCCTGCACAAAGC	646
Dd		1152	----GlyPheGlnCysThr-----CysProProAspSerProGlyProArgCys	1166
OY		647	TAAACATCCCAACACAGACAGATGTGACACAGAAC-----AGTGAAATAATGCC	694
Dd		1166	sglnArGrProGlyAlaSerGlyCysGlnGlyArgGlyValAspGlyIleThrCys--AspAl	1188
OY		695	CAAAATGTTAANAATGATGATTTACACAGCTACTGATTGGAGC-----	734
Dd		1185	agLYcys-----SerGlyProGlyGlyAspTrpSerGlyGlyAspCysSerIeucl	1200
OY		735	-----TGCCTGGCTCTTAATGCCAGGA	754
Dd		1202	yValProAspProTrpIrysgLYcysProProHisserylncIcystTrleuIlePheArgss	1222
OY		755	ATCATAG-----GGGTATGACTGCCT	775
Dd		1222	pGLyArgCysHisProGlnCysAspSerGlnGlnCysLeuPheAspGlyTyTrAspCys--	124
OY		776	CTCCAAACCCCTGG-----	788
Dd		1242	-GluIleProIeuthrCysIleProAlaTyTrAspGlnIryCysArgAspHisPheHissAs	126
OY		789	-----GGCTGTAAAGCAAGCTCAGCTAGCTCCCACTGGGGGCTGTGC	832
Dd		1261	ngLYHisCysgLYsgLYsgLYCysAsnAspAlaGlnCysglTyTrAspGlyGlyAspCysAsx	128
OY		833	CCCTCCCTGGGACGGGTCCGATGGGGAGGCCCATCACTGTTCAATAGTGAATGTA	892
Dd		1281	grProclugLYlAspSerGlnGlyIyrGrProserIeuala-----	129
OY		893	GCTAAAGCCCCCTGCTGCTGCTGCTGCTCACATGCCACAGCAGGGGGGCTGCTGGG	952
Dd		1295	-----LeuIleValIleLeuArgProPro-----AlaIleuAs	130
OY		953	ACAAATCATCGTAGACTGTCTCTCAGCTTAGTGTGCACAGAGACTTGGCGGGGATG	101
Dd		1305	pGlndIleuIleualaIeualaArgvalIleuSerIeuthrIleuArgvalGly-----	132
OY		1013	CTCCAGAGATGGGTGATTTCTGTACTCGGGAGGGCTATCTGTGACCTCCCGACAGGGGAC	107
Dd		1323	-----LeuthrVal-----ArgLysAspSerGlnGlyArgAsnMetVa	133
OY		1073	ACTCCCA-----GGCCAGCCACAGGGGTCA	109
Dd		1335	IpheProTYrProGlyThrArgAlaLysGlnGluIleuSerGlyAlaArgAspSerIerSe	135
OY		1097	GGGGCAGAGGTGCACACCCTCAGACATGAGCA-----AGAATGGGGCT	113
Dd		1355	rTrpGlnuArgGlnalaProProThcInProIeuGlnLYlsGluThcIuSerIeudGlyal	137
OY		1139	AGGG-----AGCAGGTGTGTTTATGAGCCAGGACCTGGG	117
Dd		1375	adLyPheValValaMetGlyValaAspIeuSerArgCysgLYProGlnHisProAlase	139
OY		1172	GGGG---GGGTGGGGCCGGGGCTTTCTGCTCATTTTGTCTTCAATGAAAAGCTCAAAGC	122
Dd		1395	rArgCysProTrpAspSerGlyIleuIleuIeuArgPheIeu-----AlaIleMetal	141
OY		1229	AGCCAACACAGGCTTCCCCTTCCT	1256
Dd		1412	aAlaValGlyAlaIeuGlnProIeuleu	1421

QY 429 GAGAGTAACAGGCGTGGATGTTCTTCTTGAGGCAAGACCAATTCCTCAATTGA 370
 DB 671 -----SeraspnrogluThrlLeuSerProSerThrs 681
 QY 369 CTGCCTTGAAGACAGATACCTCTCTTCATCATCATCCCTCCCTGACTTCGAT 310
 DB 681 exaspnserThrlSer-HisMetProLeuThrlSerAlaHisPro-----Thr 697
 QY 309 GGGGATCACTATCATCATCACTCCAGAGAGTCTGCCATCTCCCTTGGAGGGCTTC 250
 DB 698 GlyGlyGlu-----AsnIle-ThrGlnValThrProAlaSerIle----- 710
 QY 249 CCCACTGGCTCCTGCTCCAGGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 190
 DB 711 -----SerThrlHis 714
 QY 189 TGGCTTCACAGACATCATCAGCCTCAGCAGGTTCAGTGAAGAGAGCAGCTCCG 130
 DB 714 sValSerThrlSerSerProAlaProArgProGlyThrThr-----SerGlnAla 731
 QY 129 TGGGTAGAGTGGCCAGCAGCGGTACTCTTCCTGGG 93
 DB 731 rGlyProGlyAsnSerSerThrlThrlsProGly 743
 RESULT 13
 ID POLS RUBVM STANDARD; PRF; 992 AA.
 AC P08563;
 DT 01-AUG-1988 (Rel. 08, Last sequence update)
 DT 01-AUG-1988 (Rel. 08, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Structural polyprotein [Contains: Nucleocapsid protein C; Membrane glycoprotein E2; Membrane glycoprotein E1].
 OS Rubella virus (strain M33).
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Togaviridae;
 OC Rubivirus
 OX NCBI_TaxId=11043;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA MEDLINE=87174825; PubMed=3562245;
 RA Clarke D.M., Loc T.W., Hui I., Chong P., Gillam S.;
 RT Nucleotide sequence and in vitro expression of rubella virus 24S subgenomic messenger RNA encoding the structural proteins E1, E2 and C.
 RL Nucleic Acids Res. 15:3041-3057(1987).
 CC -1- FUNCTION: GLYCOPROTEIN E1 CONTAINS THE VIRAL HEMAGGLUTININ ACTIVITY.
 CC -1- SUBUNIT: MULTIPLE COPIES OF THE C PROTEIN COMPRISE THE NUCLEOCAPSID.
 CC -1- DISEASE: RUBELLA INFECTION DURING THE FIRST TRIMESTER OF PREGNANCY RESULTS IN DEATH OR MULTISYSTEM BIRTH DEFECTS & MICROCEPHALY.
 CC -1- THIS SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).
 CC or send an email to license@isb-sib.ch.
 DR EMBL; X05259; CAA28880.1; -
 DR PIR; A27505; GNWVR3
 KW Polyprotein; Nucleocapsid; Glycoprotein; Transmembrane; Signal.
 FT CHAIN 1 299 MEMBRANE GLYCOPROTEIN E2.
 FT SIGNAL 300 559 MEMBRANE GLYCOPROTEIN E2.
 FT CHAIN 560 580 MEMBRANE GLYCOPROTEIN E1.
 FT TRANSMEM 277 297 POTENTIAL.
 FT TRANSMEM 515 531 POTENTIAL.
 FT TRANSMEM 533 553 POTENTIAL.
 FT CARBOHYD 352 352 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 370 370 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 656 656 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 757 757 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 789 789 N-LINKED (GLCNAC. . .) (POTENTIAL).
 SQ SEQUENCE 992 AA; 106905 MW; D268889C9E78CFLE CRC64;
 Alignment Scores:
 Pred. No.: 0.444 Length: 992
 Score: 112.00 Matches: 97
 Percent Similarity: 31.99% Conservative: 38
 Best Local Similarity: 22.99% Mismatches: 129
 Query Match: 4.38% Indels: 159
 DB: 1 Gaps: 22
 US-09-989-919-15 (1-1397) x POLS RUBVM (1-992)
 QY 1184 GCCCAACCCGCCGCCAGGTCTGCTCAAC-----ACACTGCTCCCTGAGCC 1134
 DB 76 AlaProProProProGluGlnArgGlnGlnSerArgSerGlnThrProAlaProLysPro 95
 QY 1133 CAGTCTGCTCATGTGTGAGTGCACCTGCTGACCTGACCTGACCTGAGCTGGCTGGAG 1074
 DB 96 SerArg-----AlaPro-----ProGlnGlnProGln----- 104
 QY 1073 TGTCCCTGTCGGAGAGTCAGATAGCTCCCAAGTACAG-----AAT 1029
 DB 105 -----ProProArgMetGlnThrGlyArgGlySer 115
 QY 1028 CACCAACATCTGAGACATCCCGCCCAAGTCTCTGTCACAGCTAAGCTAGAGA--- 972
 DB 116 AlaProArgProGluGlnArgGlnGlnSerArgSerGlnThrProAlaProLysPro 135
 QY 971 -----ACACTCAGATGAGATTGTCACAGCAGCGCCCGCTGCTG----- 926
 DB 136 LeuArgProProLeuHisAspPro---AspThrGln-AlaProGlnGlnAlaCysValh 154
 QY 925 -----GGCATGTCAGACAGCAGCAGCAGGAGCTTACATCTCACACTAATTGAAC 870
 DB 154 rSerTrpLeuTrpSerGluGluGlyGlnGlyAlaValPheTyrArgValAspLeuHisPhe 174
 QY 869 AGTGATGGGGCTGCCCGAGAACCGTCCAGGA----- 836
 DB 174 exsnuGlyThrProProLeuAspGlnAspGlnArgTrpAspProAlaLeuMetTyrAs 194
 QY 835 -----GGGAGCAGACCCCGCAGTGGGAGAGTACAGCTGAGCTTACAGCCACAG 783
 DB 194 nProCysGlyProGlnProPro---AlaHisValValAlaGlnArgAsnGlnProAlaG 213
 QY 782 GTTGAGAGCAGATCATACCCCATGATCTGAGCTAGAGCAGCAGCTCCATAGCTA 723
 DB 213 yAspValArg-----GlyValTrpGlyGlyGlnArgThrTyr-Ala- 227
 QY 722 GCGGTGAACATCATTTTAACATT-----TGCAATTATTCAGAGTTTGTC 675
 DB 228 -----GlnGlnAspPheArgValGlyGlyThrArgTrpHisArgLeuLeuArgMetP 245
 QY 674 TGGTACATCTGCTGCTGAGATGTTAGCTTGTCTGAGAGGGGTTCCAGGTAGATGC 615
 DB 245 roVal-----ArgGlyLeuAspGlyAsp----- 252
 QY 614 AGTGTCAGTGAAGTACTCTCCACAGGCAAGCCAGCTGTAGAGATG-TCA 556
 DB 253 -----ThrAlaProLeuProProHisThrThrGlnArgGlnGlnThrArgSerA 269
 QY 555 CTGCTAGACCG---AACCTCAGATGTCAGCAGCAGCCTTACATATGTGCTCTGT 499
 DB 269 laArgHisProTrpArgGlnArgPheGlyAlaProGlnAlaPheLeuAlaGlyLeuLeu 289
 QY 498 TGGATCAGAGGACCACTGCTCATCTCAGTCTTGAAGAAAGACCACTGCGAGCTGG 439
 DB 289 euAlaAlaValAla-AlaGlyThrAlaArgAlaGlyLeuGlnProArgAlaAspMet 307
 QY 438 GGCAGAGGTGAGGTAAACAGGCGCTGATGTTCTCTTCTGAGGCAAGACCATTTCTC 379


```

KW Alternative splicing.
FT DOMAIN 297 369 RNA-BINDING (RRM) 1.
FT DOMAIN 371 452 RNA-BINDING (RRM) 2.
FT DOMAIN 9 27 3 X 3 AA REPEATS OF R-G-G.
FT REPEAT 19 21 1.
FT REPEAT 25 27 2.
FT REPEAT 25 27 3.
FT DOMAIN 10 266 GLN/GLU/PRO-RICH.
FT DOMAIN 10 15 POLY-GLY.
FT DOMAIN 20 27 POLY-GLY.
FT DOMAIN 56 65 POLY-PRO.
FT DOMAIN 67 71 POLY-GLN.
FT DOMAIN 95 98 POLY-GLN.
FT DOMAIN 99 103 POLY-PRO.
FT DOMAIN 184 188 POLY-PRO.
FT DOMAIN 571 574 POLY-ARG.
FT DOMAIN 613 616 POLY-GLY.
FT DOMAIN 635 641 POLY-GLY.
FT VARSPLIC 663 707 PTERFPGAGCPVGGGPRGPGPTAGYGRGREYEG
FT CONFLICT 243 243 PNRKPRF -> VRMIDVG (IN SHORT ISOFORM).
SQ SEQUENCE 707 AA; 76149 MW; 68D5DEA5E235847 CRC64;
G -> R (IN REF. 3).

Alignment Scores:
Pred. NO.: 0.51 Length: 707
Score: 111.00 Matches: 49
Percent Similarity: 36.75% Conservative: 12
Best Local Similarity: 29.52% Mismatches: 70
Query Match: 4.34% Indels: 35
Db: 1 Gaps: 7

US-09-989-919-15 (1-1397) x SFPQ_HUMAN (1-707)
QY 1192 AGGCCCCGAGCCCCAGCCCCGAGGCTGCTGACTCAACACACACACTGCTCCTGACCCC 11333
Db 86 GlnProProProHisProGlnProHisGlnGlnGlnGlnProPro-----ProProPro 103
QY 1132 AGTTTGACTCATGCTGAGGTGTCACCTTGCCCTGACCCCTGGGCTGGGCTGGAGGT 10737
Db 104 GlnApsPserSer-LysProValValAlaGlnGlyProGlyProAlaProGlyValGlySe 123
QY 1072 GTCCCTGTGGGGAGGTCAAGATAGCCTCCCGCAGTACAGATACACCCCATCTCTGGAG 10133
Db 123 rAlaProProAlaSerSer-----SerAlaProProAlaThrProProThrSerGlyAl 141
QY 1012 CATCCCGCCGCAAGTCTCTGTGCCAGACTAAGCTGAGAGAGAACTCCACGATGGATTGT 953
Db 141 aProProGly-----SerGlyProGly----- 148
QY 952 CCCGACGAGCCCCCGACCGCTGCTGTGGCATGTGCACAGACAGCAGGAGGCTTTAGC 893
Db 149 -ProThrProThrProProProAlaValThrSerAlaProProGlyAlaProProProTh 168
QY 892 TAACTTTCTACACATTTGAACACACAGTATGGGGCTGCCCGCAGGAACCTCCAGGAGAG 833
Db 168 rProProSerSer-----GlyValProThrThrProProGlnAlaG1 182
QY 832 GCACAGAGCCCCAG-----TGGGAGACTAGCCTGAGCTTGCTTA 794
Db 182 yGlyProProProProProAlaAlaValProGlyProGlyProGlyProGlnGlyPr 202
QY 793 CAGCCCAAGAGGTTGAGAGGAGCAGTCAATCCCATGATTCCT-----GGACT 746
Db 202 oGlyProGlyGlyProLysGlyGlyLysMetProGlyGlyProLysProGlyGlyGlyPr 222
QY 745 AGGAGCCAGCAGATGCC 730
Db 222 oGlyLeuSerThrPro 227

```